

Lincoln University Digital Dissertation

Copyright Statement

The digital copy of this dissertation is protected by the Copyright Act 1994 (New Zealand).

This dissertation may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- you will use the copy only for the purposes of research or private study
- you will recognise the author's right to be identified as the author of the dissertation and due acknowledgement will be made to the author where appropriate
- you will obtain the author's permission before publishing any material from the dissertation.

Polymorphism in the human Cathepsin B gene

A dissertation
submitted in partial fulfilment
of the requirements for the degree of
Bachelor of Science with Honours
at Lincoln University

Jason R MacKenzie

November 1995

Animal and Veterinary Sciences Group

Lincoln University,
Canterbury, New Zealand.

Abstract

Cathepsin B is a cysteine protease which has been implicated in the development and progression of various cancers. It is often redistributed to the plasma membrane, and inactive forms of the enzyme are commonly found in tumour tissue exudates. The secretion of cathepsin B from cancer cells is thought to correlate with their invasive potential. The purpose of this study was to identify and characterise polymorphism in the human cathepsin B gene.

PCR was used in the amplification of exon 7 and intron 7 of the human cathepsin B gene, using primers designed from the reported bovine and human cathepsin B genes. The two resulting amplicons were of different size, revealing a potential polymorphism.

10 human volunteers were screened for this variation, and sequencing of the variant amplicons identified two previously unreported alleles of the human cathepsin B gene. One allele is thought to confer increased stability to the primary transcript for cathepsin B, and hence may alter gene expression in homozygous carriers.

Acknowledgments

Sue Mason, many thanks to you for your patience and understanding throughout the year. Your guidance has been much appreciated. Thanks also for the time you have put into my project. Also to Jon Hickford, thank you for your time, your effort, and your critical appraisal when it was needed. I'm sure they make red ink just for you.

A special thank you to SM, KT, JH, BP, JB, AC, JP, MD, LV, & ME. This work would have been nothing without you (especially ME).

Thank you Roy Bickerstaffe, for the opportunity to undertake this work. I know I have gained invaluable experience and skills whilst working in your laboratory. A word of advice for Roy's future students - smile often, it makes him wonder what you're up to.

Noelle Roberts, many thanks for being so helpful and resourceful. Thanks for keeping my bench tidy too, I never could find anything after we had important visitors.

A very special thank you to Kate and Matthew, who set aside the time to help me with all the small but very important things.

My parents: Dad, Ali, Mum, and John. Thank you all for your understanding. I will try to get home more often! Thanks also for the computer, it has been my right hand this year.

Once again, my special gratitude to the Sunday editors - you know who you are.

Contents

Table of contents	Page
Title	i
Abstract	ii
Acknowledgments	iii
Table of contents	iv
Table of figures	vii
Table of tables	vii
List of abbreviations	viii
 Chapter 1 Introduction	
1.1 Cathepsins	1
1.2 Properties of cathepsin B	2
1.3 Cathepsin B and cancer	3
1.4 Cysteine protease inhibitors and cancer	5
1.5 The human cathepsin B gene	6
1.6 Evidence for alternative splicing of cathepsin B transcripts	7
1.7 Polymorphisms in the cathepsin B gene	10
1.8 Objectives	11
 Chapter 2 Amplification of the human cathepsin B gene	
2.1 Introduction	12
2.1.1 Selection of primers and development of initial PCR reaction	12
2.1.2 Development of whole blood PCR protocol	13
2.1.3 Agrose gel electrophoresis and visualisation of amplimers	14
2.1.4 Revised whole blood PCR	15
2.1.5 Discussion	16
2.2. Refinement of PCR methodology	16
2.2.1 Introduction	16
2.2.2 Results	17

2.2.3 Discussion	18
Chapter 3 Preparation of 10 human genomic DNA samples for gene analysis	
3.1 Introduction	19
3.2 Hi-Salt DNA extraction from whole blood	19
3.3 Estimation of raw genomic DNA concentrations	20
3.4 Results and discussion	21
Chapter 4 PCR amplification of genomic DNA from 10 individuals	
4.1 Introduction	22
4.2 PCR reagents and reaction protocol	22
4.3 Results and discussion	22
Chapter 5 Preparation of purified amplimers for sequencing	
5.1 Introduction	24
5.1.1 Wizard TM PCR preps DNA purification system (Promega)	24
5.1.2 Agrose gel DNA extraction kit (Boehringer Mannheim)	24
5.1.3 Results and discussion	25
5.2 Measurement of purified amplimer concentration by gel electrophoresis	26
5.2.1 Results and discussion	26
5.3 Estimation of DNA concentrations using the scanning spectrophotometer	27
5.3.1 Results and discussion	27
Chapter 6 Sequencing results	
6.1 Introduction	28
6.2 Sequence failure due to errors in estimation of DNA concentration	28
6.3 Sequence results from the University of Auckland	29
Chapter 7 Discussion of sequence results	
7.1 Introduction	30
7.2 Intron consensus sequences within alleles A and B	30
7.3 Implications of the extra 19 base pairs in allele A	31

7.4 Identification of possible point mutation	34
Chapter 8 Conclusions	
8.1 Conclusions	36
8.2 Future research in this area	36
References	37
Appendix A.1 Primer homology results from the NCBI blast serv	40
Appendix A.2 Visualisation of DNA by gel electrophoresis	44
Appendix B.1 Consent form	46
Appendix B.2 Hi-salt DNA extraction reagents	47
Appendix C.1 Homology between A and B alleles	48
Appendix C.2 Sequence signal strengths	49
Appendix C.3 19 nucleotide sequence search results	53

Table of figures

	Page
1.1 Interaction between cathepsin B and stefin B	6
1.2 Structural organisation of the human cathepsin B gene	7
1.3 Cathepsin B transcripts in human tissues	9
2.1 Primer alignment with human cathepsin B gene sequence	13
2.2 Agrose gel of whole blood PCR	15
2.3 PCR banding pattern for subject 4 over a range of magnesium ion concentrations	17
2.4 PCR banding patterns generated from DNA of human Subject 4 and tumour tissue	18
3.1 Scan of pure DNA sample XXX	20
4.2 Banding pattern suggestive of two different alleles amongst ten human subjects	23
5.1 Purified amplimers on agrose gel for concentration estimation by band density	27
6.1 Consensus sequences for amplimers 10 and 8	29
7.1 Region of possible point mutation	30
7.2 General consensus sequence for introns	31
7.3 The stable loop structure afforded by 19 extra nucleotides in allele A	32
7.4 The stable loop structure formed in absence of the 19 extra nucleotides in allele A	32
7.5 Signal strengths for the regions containing possible point mutations	34

Table of tables

	Page
1.1 Active site region amino acid homologies among papain-type cysteine residues	1
1.2 Nucleotides at the intron-extron boundaries of the human cathepsin B gene	8
3.1 Raw genomic DNA concentrations and purity	21
7.1 Probable branch sites for the A and B alleles	31
7.2 The possible alternative branch site in allele A	33
7.3 Possible point mutation between alleles A and B	34

List of Abbreviations

12Py	12 pyrimidine
A	adenine
aa	amino acid
A ₂₆₀	absorbance at 260nm
A ₂₈₀	absorbance at 280nm
bp	base pairs
C	cytosine
cDNA	complementary DNA
conc	concentration
DNA	deoxyribonucleic acid
dNTP	2'-deoxyribonucleic triphosphate
EDTA	(ethylenedinitrilo) tetraacetic acid disodium salt
g	gram
G	guanine
kcal mol ⁻¹	kilocalories per mole
kb	kilo base pairs
kDa	kilodaltons
KCl	potassium chloride
Leu	leucine
M	molar
mg	milligram
mM	millimolar
mol	mole
mRNA	messenger ribonucleic acid
NCBI	National Centre for Biotechnology Information
ng	nanograms
nm	nanometre
PCR	polymerase chain reaction
pmol	picomole
RNA	ribonucleic acid

r.p.m	revolutions per minute
T _m	melting temperature
T	thymine
μL	microlitre
μg	microgram
UV	ultraviolet
V	volt
Val	valine
[]	concentration

CHAPTER 1

Introduction

1.1 Cathepsins

Cathepsins are members of a large family of papain-type cysteine proteases, which also includes the calcium dependent calpains. Unlike the cytosolic calpains, the cathepsins are located in cell lysosomes, where they are predominantly found in active form. Cathepsins B, H, L, and S are all closely related through active site sequence homology and are similar in size, with the mature forms ranging from 24-32 kDa (Chapman *et al.*, 1994). While they share high active site homology, the overall amino acid homology among the cathepsins is much less striking (20-60%). The two closest cathepsins in terms of sequence similarity are cathepsins L and S, which share 60% homology (Chapman *et al.*, 1994). Table 1.1 illustrates conserved active site amino acid sequences between papain and the cathepsins.

**Table 1.1 Active site region amino acid homologies
among papain-type cysteine proteases (from Chapman *et al.*, 1994).**

	19	*	170
Papain	--NQGSCGSCWAFS---		YILIKNSWG
	19	*	179
Cathepsin S	--YQGSCGACWAFS---		YWLVKNSWG
	19	*	182
Cathepsin L	--NQGQCGSCWAFS---		YWLVANSWN
	23	*	214
Cathepsin B	--DQGSCGSCWAFG---		YWLVANSWN
	20	*	182
Cathepsin H	--NQGACGSCWTFS---		YWLVKNSWG

* denotes the active site cysteine residue

The cathepsin precursors all have *N*-linked oligosaccharides containing mannose- 6-phosphate, which is typical of proteins that accumulate in lysosomes (Moin *et al.*, 1992; Chapman *et al.*, 1994; Mach *et al.*, 1994a). The lysosomal cathepsins are accepted as the proteases which

degrade defective or damaged proteins within the lysosome, and mediate the recycling of amino acids for new protein synthesis (Moin *et al.*, 1992; Mordier *et al.*, 1995). They are optimally active at acidic pH - especially human cathepsin B, which suffers irreversible loss in activity and changes to secondary structure when exposed to pH ≥ 7.0 (Turk *et al.*, 1994). Therefore the containment of cathepsins within the cell lysosome can be regarded as a protective mechanism. Any cathepsin activity accidentally released into the cell cytosol will not indiscriminately digest protein substrates, but is inactivated by the higher pH in the cell cytosol.

1.2 Properties of cathepsin B

Cathepsin B is the only cathepsin to possess C-terminal dipeptidase (endo- and exopeptidase) activity (Chapman *et al.*, 1994; Zongchao *et al.*, 1995). The mature protein exists in two forms: a single chain form of ~30 kDa (254 residues) and a double chain form with a heavy 25 kDa chain and a lighter 5 kDa chain (Mordier *et al.*, 1995). The double chain form is produced by enzymatic cleavage between residues 47 and 50, with the loss of a dipeptide (Moin *et al.*, 1992; Mordier *et al.*, 1995). Human cathepsin B contains six disulphide bridges, one of which links the heavy and light chains (Mordier *et al.*, 1995).

Like other cathepsins, cathepsin B is synthesised as a latent precursor, or preproenzyme which is subsequently converted to its mature form by limited proteolysis. All mammalian species produce a preproenzyme containing a 17-residue prepeptide, a 62-residue proregion, a 253-residue mature proteinase (254 in bovine), and a C-terminal extension of 6 residues for rat, mouse and human, and three residues in bovine cathepsin B (Mordier *et al.*, 1995). Procathepsin B is cotranslationally glycosylated at Asn 113 (numbered according to the mature enzyme), and possibly at Asn 42 in the proregion. A third potential site is found in human cathepsin B at Asn 120. Phosphorylation of mannose residues in *cis*-Golgi allows for the specific binding of procathepsin B by mannose 6-phosphate receptors in the *trans*-Golgi network, to ensure accurate sorting to prelysosomal compartments (Mordier *et al.*, 1995).

Activated cathepsin B is predominantly found in lysosomes, and it was originally assumed that this was the site of activation. Activated forms of cathepsin B have been found elsewhere within cells (endosomal compartments) suggesting that processing may occur earlier in the biosynthetic pathway, and is not strictly limited to the lysosome. It is now thought that the proteolytic processing of cathepsin B is initiated in prelysosomal compartments, and involves the cleavage of the propeptide, and the C-terminal extension. Excision of the dipeptide (residues 48 and 49) may occur in the lysosome through the action of other resident proteases. Removal of the dipeptide within the lysosome may decrease the stability of the active protein in alkaline conditions, and ensure that any activity lost to the cell cytosol is quickly inactivated. Mach *et al.*, (1994a) have reported that recombinant procathepsin B can be activated *in vitro* by a unimolecular autocatalytic mechanism, and that processing of the activated procathepsin to the mature form is a result of intramolecular processing. It is unknown as to what extent the proenzyme undergoes self activation and processing to the mature form *in vivo*.

1.3 Cathepsin B and cancer

The expression, post translational processing, and targeting of cathepsin B are all frequently altered in transformed and malignant cells (Gong *et al.*, 1993; Calkins and Sloane, 1995; Mordier *et al.*, 1995). Increased expression or activity of cathepsin B has been observed in a variety of both human and murine tumours, specifically at the invasive edges of human prostate carcinoma (Cao *et al.*, 1994). Altered forms of cathepsin B (differing in size and stability) have been found in a variety of animal tumour cells and tissue culture exudates, including fluids from human breast and ovarian cancers. The elevated levels and variation in both form and proteolytic potency of this enzyme has led to its implication as a potentiator of tumour invasiveness and metastasis (Moin *et al.*, 1992; Gong *et al.*, 1993).

The peripheral distribution of cathepsin B protein and activity explains the cosedimentation of cathepsin B with plasma membrane/endosomal fractions of malignant human and murine cells (Moin *et al.*, 1992). Therefore levels of membrane associated cathepsin B may indicate the

metastatic or invasive potential of these tumours, due to its altered distribution and its ability to degrade extracellular matrix and basement membrane proteins (Turk *et al.*, 1995). The transfection of MCF-10 human breast epithelial cells with oncogene *ras* leads to the association of cathepsin B containing lysosomes with the cell membrane (Sloane *et al.*, 1994). The ability of oncogene *ras* to affect intracellular trafficking of cathepsin B is intriguing, as the re-distribution of lysosomes from around the nucleus to the cell periphery is a phenomena usually reserved for cells involved in degradative or invasive processes.

Although cathepsin B is irreversibly inactivated in alkaline pH (Turk *et al.*, 1994), active, high molecular mass forms have been reported in cultured tumour tissue exudates. Mach *et al.*, (1994b) observed the transient formation of a complex between the mature enzyme and its propeptide during autocatalytic proteolytic maturation of recombinant procathepsin B. They suggest that this complex occurs *in vivo*, and accounts for the high molecular mass forms present in tumour tissue exudates. Extracellular high molecular mass forms of cathepsin B may serve as latent enzyme pools, and dissociation of the complex followed by degradation of the inhibitory propeptide may occur in the acidic micro-environment surrounding living cells, *i.e.* in the extracellular matrix. The mature enzyme may then act on potential protein substrates until it becomes inactivated by the secretion of cysteine protease inhibitors, or by the neutral pH commonly encountered in body fluids (Mach *et al.*, 1994b).

Elevated cathepsin B gene expression is manifest through increased levels of mRNA and protein, during the early development of colorectal cancer (Murnane *et al.*, 1991; Shuja *et al.*, 1991). Levels of cathepsin B activity in tumours associated with regional lymph-node metastasis, distant metastasis, and in normal tissues are significantly lower than levels in tumours which are confined to, or are invading the bowel wall. Therefore increased cathepsin B gene expression is a characteristic of tumours actively invading the bowel wall or local tissues (Murnane *et al.*, 1991), and a decrease in gene expression seems to occur as the disease progresses. mRNA transcripts of different size (2.2, 4.0, 1.5 and 3.0 kb) are also evident, and may reflect alternative splicing of the primary transcripts.

Increased levels of cathepsin B activity in serum, and in tumour tissue are emerging as a prognostic factor in cervical carcinoma. Makarewicz *et al.*, (1995) measured cathepsin B activity levels in 106 patients with cervical carcinoma, to estimate the usefulness of cathepsin B activities in determining the extent of the disease. They observed a positive correlation between serum cathepsin B levels and levels of cathepsin B in tumour tissue. In contrast to colorectal cancer, increased levels of tumour cathepsin B activity correlated with tumour metastases to pelvic lymph nodes in patients. Makarewicz *et al.*, (1995) concluded that levels of cathepsin B activity may be useful in predicting the extent of cervical carcinoma in clinical practice.

1.4 Cysteine protease inhibitors and cancer

Cysteine protease inhibitors (CPIs) are pseudoirreversible inhibitors with a high affinity for cathepsins (Calkins & Sloane, 1995). They contain a highly conserved amino acid sequence - QVVAG-, which is partly responsible for their tight binding. Endogenous CPIs constitute a single protein superfamily called the cystatins, which can be divided into four families: stefins (family I), cystatins (family II), kininogens (family III), and some non-inhibitory proteins (family IV).

The cystatins are proposed to bind with cysteine proteases in areas adjacent to the active site, and exert their inhibitory effects by physically blocking the active site. There is no specific interaction with the cysteine, or any other residues within the active site. Figure 1.1 illustrates how stefin B may interact with cathepsin B.

Stefin B is a more potent inhibitor of cathepsins L and S, and Figure 1.1 suggests that an equilibrium exists between stefin B and cathepsin B. The region of stefin B which occupies or blocks the active site is referred to as the "trunk".

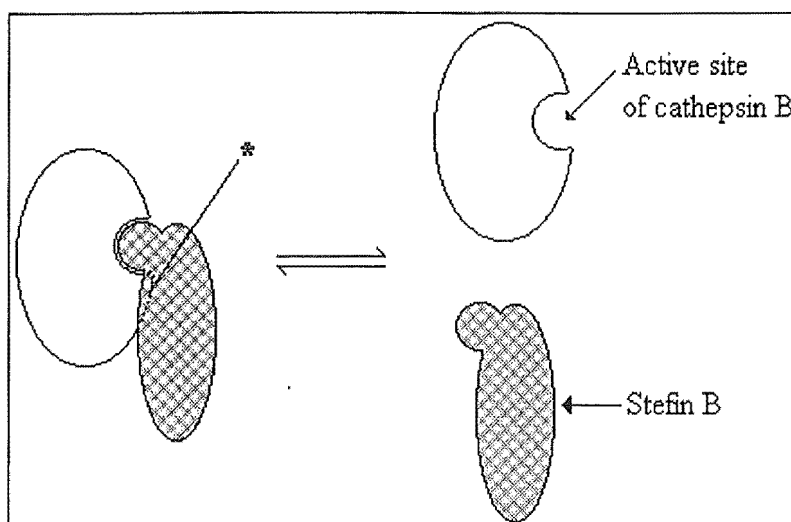


Figure 1.1 Interaction between cathepsin B and stefin B. "*" represents the region containing the -QVVAG- consensus sequence (adapted from Calkins & Sloane, 1995).

The progression of malignant tumours may be due in part to an imbalance in the CPI : cysteine protease ratio. A decrease in CPI activity or concentration may contribute to this imbalance. Two out of three breast carcinomas are reported to contain low CPI activity in comparison to normal breast tissue (Calkins and Sloane, 1995). As a result, breast cancer tissues possess increased cathepsin B and L activity which is probably due to reduced CPI activity. The remaining third of breast carcinomas have CPI activity equal to or higher than normal breast tissue. Stefin A in particular has come under scrutiny, as 4 out of 5 breast carcinomas have decreased levels of stefin A mRNA transcript (Lah *et al.*, 1992; cited in Calkins and Sloane, 1995). Reductions in stefin A protein levels parallel the decrease in CPI activity and stefin A mRNA, suggesting stefin A is a major contributor to the observed CPI activity in breast tissue (Calkins & Sloane, 1995). Future study in this area will elucidate the involvement of cystatins in tumour progression.

1.5 The human cathepsin B gene

There is increasing evidence for the existence of one gene for human cathepsin B, despite the presence and generation of multiple mRNAs in both normal and transformed tissue. The

cathepsin B gene had previously been mapped to two locations: 8p22 and 13q14, but has since been assigned to chromosome 8 at 8p22-p23.1 (Fong *et al.*, 1992). The gene is large (in excess of 18 kb), and contains 12 exons (Fig 1.2). Parts of exons 3 and 11, and the intervening exons (4-10) are translated, and constitute preprocathepsin B.

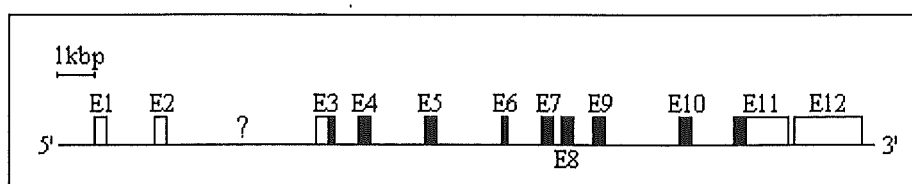


Figure 1.2 Structural organisation of the human cathepsin B gene. Solid areas are regions that encode preprocathepsin B. The sequence is deduced from two clones which contained discontinuous segments of the human cathepsin B gene, and "?" represents the unknown length of intron between exons E2 and E3 (adapted from Gong *et al.*, 1993).

In mammals, the cathepsin B gene exhibits a complete conservation of the intron exon junctions in the coding portion of its gene. The size and sequence of introns diverge between species (Mordier *et al.*, 1995). Human cathepsin B is encoded by exons 3-11 (Gong *et al.*, 1993), and Table 1.2 shows the nucleotides at the intron-exon boundaries, and the size of introns in the human cathepsin B gene.

1.6 Evidence for alternative splicing of cathepsin B transcripts

Different forms of cathepsin B and multiple mRNA types for this enzyme are produced in both normal and human tumour tissues, and generated early speculation that more than one cathepsin B gene existed. For example both normal and cancerous colorectal mucosa contain 2.2 and 4.0 kb mRNA transcripts for human cathepsin B, whereas human osteoclastomas can contain 2.4, 1.9, and 1.2 kb cathepsin B related mRNAs. Gong *et al.*, (1993) isolated and characterised cDNA and genomic clones encoding cathepsin B from several human tissues and tumours. Human kidney and HepG2 (hepatocyte) cells were found to contain at least four different types

of mRNA for cathepsin B, and a fifth type of transcript was present in only some human tumours. The preprocathepsin B coding sequence in all of the five different transcripts was identical, suggesting that the different mRNAs were derived by alternative splicing of mRNA derived from a single cathepsin B gene (Gong *et al.*, 1993).

Table 1.2 Nucleotides at the intron exon boundaries of the human cathepsin B gene (from Gong *et al.*, (1993)).

Exon	Exon size (bp)	5' splice donor	Intron size (bp)	3' splice acceptor
		5' UTR		5' UTR
1	124	CGCTGGgtgag...	~3500	...tcccagGCTGGT
		5' UTR		5' UTR
2	88	CGAGTAgtag...	>5000	...ctccagGTGGAT
		TrpGln		AlaGly
3	151	TGGCAGgtagg...	820	...ttccagGCCGGG
		GlnArg		ValMet
4	86	CCAGACgtgag...	1450	...atccagAGTTATG
		CysTrp		AlaPhe
5	115	TGCTGGgttaag...	1750	...ctgcagGCCTTC
		AspGly		CysAsn
6	119	GGACGGgtaag...	750	...tttcagCTGTAAT
		HisVal		GlyCys
7	86	CATGTAGgtaag...	230	...catcagGGTGCA
		HisTyr		GlyTyr
8	144	CACTACGgtaag...	470	...ttccagGATACA
		LysSer		GlyVal
9	117	AAGTCAGgtgag...	1380	...tttcagGAGTGT
		AspAsn		GlyPhe
10	129	GACAATGgtgag...	470	...tcccagGCTTCT
		3'UTR		3' UTR
11	142	ATCGGGgtaga...	823	...ctgcagAGAATC
12	2700			

The five different transcripts characterised by Gong *et al.*, (1993) are illustrated in Figure 1.3. Type 1 is the largest transcript produced, and contains a 237 nucleotide 5' untranslated region (UTR), 1017 nucleotides which encode preprocathepsin B, and a 2800 nucleotide 3' UTR. Types 1 and 2 (4 kb in length) differ from each other through the deletion of 88 nucleotides in the 5' UTR in type 2. Types 3 and 4 (2.3 kb in length) differ from each through the same deletion of 88 nucleotides in the 5'UTR in type 4.

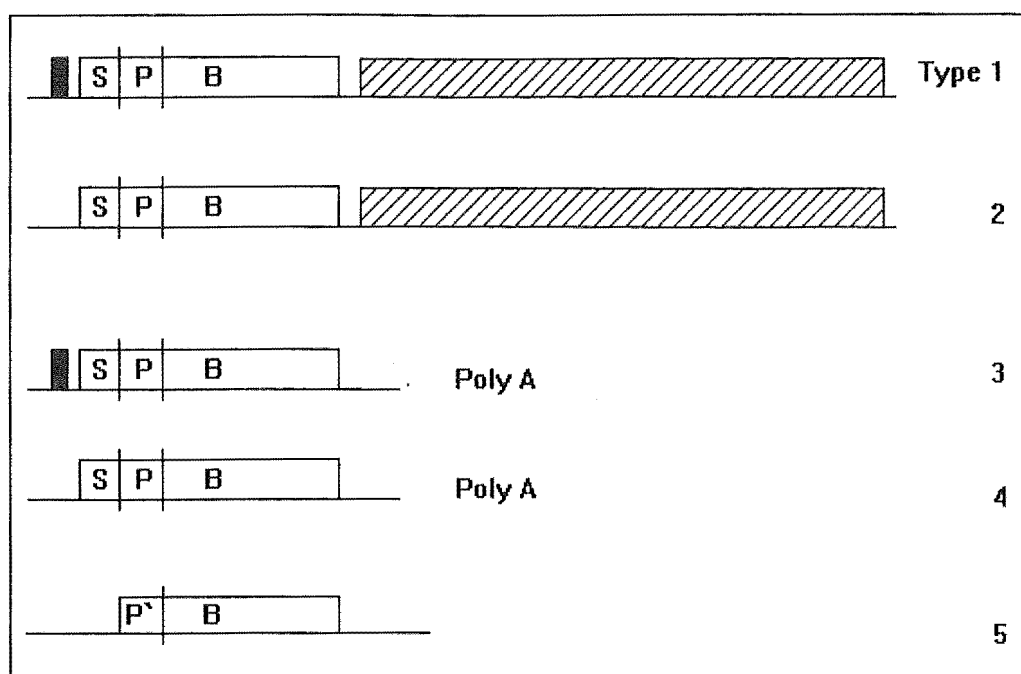


Figure 1.3 Cathepsin B transcripts in human tissues. Overlapping cDNA clones corresponding to mRNA types 1, 2, 3, and 4 were isolated from human kidney and HepG2 phage libraries. The type 5 transcript was characterised by PCR, and the 3' untranslated region associated with this transcript has not been determined. Solid black boxes indicate the 88 nucleotide sequence contained in mRNA types 1 and 2, but deleted in mRNA types 3, 4, & 5. The type 5 transcript also has an additional 155 nucleotide deletion corresponding to the signal peptide (S), and 7 amino acid residues of the propeptide (P). Cross hatched boxes indicate the alternative 3'- UTRs (adapted from Gong *et al.*, 1993).

Processing at a cryptic intron donor site in exon 11 and splicing to exon 12 produces the larger type 1 and 2 mRNAs, with alternate 3' UTRs. mRNAs with the 88 nucleotide deletion in their 5' UTRs are produced as a result of the variable removal of exon 2 (types 3, 4 and 5).

Gong *et al.*, (1993) observed that the ratio of 4.0 kb to 2.3 kb mRNA transcripts in most tissues examined was 2:1, but the ratio of 5' UTR variants differed widely. In particular, mRNAs lacking exon 2 were most predominant in human tumours. Human melanoma, breast and colon carcinomas all contained the type 5 transcript which lacks exons 2 and 3 which code for the signal peptide and 7 amino acids of the activation propeptide.

An *in vitro* transcription/translation assay was performed by Gong *et al.*, (1993), and demonstrated that the different mRNAs had different rates of translation. The relative rates of translation for the type 5 transcript, types 3 and 4, and types 1 and 2 were 8:2:1, and indicate that

the smallest type 5 transcript can be translated *in vitro*, and quite quickly in comparison to mRNAs which contain the full 5' UTR (types 1 and 3). Gong *et al.*, (1993) concluded that the expression of cathepsin B in human tissues may be regulated in part at the level of mRNA processing.

In vitro translation of the type 5 transcript is thought to start from an in-frame methionine residue (52) within the propeptide region. The resulting 32 kDa product lacks the signal peptide and the first 34 residues of the 62 amino acid propeptide region. It is no longer targetable to the lysosome, and could explain why some tumour tissues show a redistribution of cathepsin B to the plasma membrane, secretory vesicles, and to the cell nucleus. Tumour cells are known to secrete higher molecular weight forms of cathepsin B, in latent and active forms. This could reflect alterations in the linked targeting and post-translational processing of cathepsin B that result in changes in the amount and composition of the oligosaccharide side chains, and the extent of proteolytic modifications. The findings of Gong *et al.*, (1993) also imply that variable splicing of the cathepsin B mRNA transcript can also complement the diversity of cathepsin B form and function seen in both normal human and tumour tissues. Translation of the type 5 transcript *in vivo* has yet to be determined, but if translation occurs, then the protein product could be located in the cell cytosol and/or the nucleus. Its role could be in the modulation of cellular activity through the proteolytic modification of key regulatory proteins (Gong *et al.*, 1993).

1.7 Polymorphisms in the cathepsin B gene

There are few identified polymorphisms in the human cathepsin B gene. Those which have been published are usually identified through the use of complimentary DNA sequences. Variation in the 5' and 3' UTRs is often seen, but has little effect upon the size and shape of the protein produced.

Cao *et al.*, (1994) analysed full length human preprocathepsin B cDNA clones from gastric adenocarcinoma for sequence modifications potentially linked to the altered intracellular

trafficking and secretion of cathepsin B. A 10 bp insertion (CATCAGACCT) in the 3' UTR not found in the kidney and hepatoma cathepsin B cDNAs by Chan *et al.*, (1986) was identified. Cao *et al.*, (1994) concluded that the 10 bp insertion could potentially alter mRNA stability and hence alter cathepsin B gene expression. Tam *et al.*, (1994) independently characterised the same 10 bp insertion in the 3' UTR of cathepsin B in lung fibroblasts. With reference to the results of Cao *et al.*, (1994) Tam *et al.*, (1994) postulated that the 10 bp insertion occurs naturally, and may have been deleted in the cDNAs studied by Chan *et al.*, (1986). The 10 bp insertion can take part in the formation of a stable stem loop structure within the cathepsin B mRNA transcript, and may provide binding sites for regulatory proteins that can modify the stability of this message. If the 10 bp insertion was deleted, then the kidney and hepatoma cathepsin B mRNAs characterised by Chan *et al.*, (1986) may have had altered stability in comparison to transcripts from fibroblasts, or from the human gastric adenocarcinoma.

In addition to the 10 bp insertion, Cao *et al.*, (1994) also reported the following variations in the human cathepsin B gene, which were identified using four cDNA clones isolated from a human gastric adenocarcinoma cDNA library:

- a nucleotide substitution in the coding region for the propeptide from GTG to CTG, resulting in a Val²⁶⇒Leu change.
- three silent nucleotides replacements in the coding region for the mature protein.
- five single nucleotide substitutions in the 3' and 5' untranslated regions.
- heterogeneity in the 5' untranslated region.
- two restriction fragment length polymorphisms *Eco* RI and *Taq* I.

1.8 Objectives

The objectives of this study were to identify any polymorphism(s) in the human cathepsin B gene using PCR technology, and to characterise any polymorphism(s) found.

CHAPTER 2

Amplification of the human cathepsin B gene.

2.1 Introduction

The Polymerase Chain Reaction (PCR) was developed in 1985, and enables users to amplify specific gene sequences from a complex mixture of genomic material (PCR Protocols, 1990). After amplification the amplimers can be visualised as bands on agarose gels.

The polymerase chain reaction was used in this study to amplify what would potentially be 342 nucleotides of the human cathepsin B gene, including intron 7. This would be accomplished using primers that shared 100% homology with human cathepsin B gene sequence published by Chan *et al.*, (1986).

2.1.1 Selection of primers and development of initial PCR reaction

The selection of adequate primers is a critical step in the development of a PCR method. In a previous study (Markham, 1994), two primers were designed to amplify a region of the ovine cathepsin B gene. They were: Bocthex8 (5'-GGGAGGGATGGAGTACGGTCTGCA-3'), and BovcthU (5'-GAGTGGCTCTTTTCTCCAGCTGTAACGG-3'). Bocthex8 is located in exon 8 of the human cathepsin B gene, and also shares 100% homology with bovine cathepsin B sequence (Appendix A.1). The BovcthV primer was originally used by Mordier *et al.*, (1993) to study polymorphism in the bovine cathepsin B gene.

In this study, Bocthex8 and BovcthU were used to try and amplify human cathepsin B. No amplification of any part of the human genome was observed (results not shown).

Gong *et al.*, (1993) reported the nucleotides surrounding the intron-exon boundaries for human cathepsin B, and also reported the size of each intron and exon (Table 1.2 pp 8). Several introns were sufficiently small enough to be used as potential candidates for PCR amplification. The previously used primer Bocthex8 was complementary to the start of exon 8 (Fig 2.1) of the human cathepsin B gene sequence first published by Chan *et al.*, (1986). Due to the availability of a suitable primer in exon 8 and the small size of intron 7 (230 bp), a region spanning intron 7 was chosen for amplification using a new primer located upstream in exon 7. The computer Oligo™ computer program was used to identify a suitable primer in the 5' end of exon 7 which formed limited primer-dimer with itself and Bocthex8, and limited stable secondary structure such as loops, which can interfere with the PCR reaction.

This primer (named "HucathU"), is shown in Figure 2.1. When used in conjunction with Bocthex8, the 342 bp amplicon was anticipated to contain most of exon 7 (85 of 86 bp), all of intron 7 (230 bp), and the first 27 bp of exon 8.

```

HucathU      TGTAATGGTGGCTATCCTGCTGAAG
Exon 7      CTGTAATGGTGGCTATCCTGCTGAAGCTTGGAACTTCTGGACAAGAAAAGGCCT
            GGTTTCTGGTGGCCTCTATGAATCCCATGTA

            -----intron7-----

Bocthex8     ACGTCTGGCATGAGGTAGGGAGGG
Exon 8      GGGTGCAGACCGTACTCCATCCCTCCCTGTGAGCACCACGTCAACGGCTCCCGG
            CCCCCATGCACGGGGGAGGGAGATACCCCAAGTGTAGCAAGATCTGTGAGCCT
            GGCTACAGCCCGACCTACAAACAGGACAAGCACTACG

```

Figure 2.1 Primer Alignment with Human Cathepsin B Gene Sequence (Gong *et al.*, 1993; Chan *et al.*, 1986).

G + C content was 62.5% for Bocthex8, and 48% for HucathU. The melting temperature (T_m) was calculated for each primer using the methods of Thein & Wallace (1986), giving $T_m = 78^\circ\text{C}$ for Bocthex8 and $T_m = 74^\circ\text{C}$ for HucathU.

Homology of both HucathU and Bocthex8 primers with the human cathepsin B gene was confirmed using the NCBI Blast server (Blast@NCBI.NLM.NIH.GOV). The results of each search are in Appendix A.1. Synthesis of the HucathU primer was performed by Canterbury Health Laboratories, Christchurch, New Zealand.

2.1.2 Development of whole blood PCR protocol

Ohhara *et al.*, (1994) previously reported the successful amplification of a mouse retinoblastoma susceptibility gene from DNA in microwave-irradiated whole blood. It was concluded that microwave irradiation destroyed cell structures and exposed the genomic DNA, and that it also denatured proteins or proteases which could interfere with the PCR reaction.

Experimentation with microwave-irradiated whole blood PCR was carried out. Subject 1 volunteered 2 μL aliquots of whole blood (obtained with a lancet), which were placed in a PCR reaction tube and microwaved on high over a range of times from 1 - 4 minutes. The microwaved blood then had the following PCR reagents added on ice:

10 x PCR reaction buffer (Boehringer Mannheim)	2.0 μL
Water (Appendix 2.A)	14.8 μL
HucathU (50 mM)	0.4 μL
Bocthex8 (50 mM)	0.4 μL
dNTPs (0.25 mM) (Boehringer Mannheim)	0.2 μL
TAQ polymerase (50 U/ μL) (Boehringer Mannheim)	0.2 μL
Total volume (including blood)	20.0 μL

A reagent blank containing PCR reagents and 2 μL of water instead of whole blood was included during every PCR amplification, and all tubes were overlaid with mineral oil to inhibit evaporation. PCR amplification was carried in a COY™ Tempcycler II multicycler (COY Laboratory Products Inc.) using the following conditions recommended by Couldrey (1995). Tubes were subjected to an initial pre-denaturation step of 94°C for 3 minutes. This was followed by 35 cycles of denaturation at 94°C for 30 seconds, and annealing and primer extension at 60°C for 6 minutes. Tubes were stored at 4°C overnight if necessary. PCR products were viewed following gel electrophoresis.

2.1.3 Agarose gel electrophoresis and visualisation of amplimers

Agarose gel electrophoresis is a quick and simple way to determine both the size and quantity of PCR products, and can also be used to check primer specificity. If binding is non-specific (two or more bands of DNA are produced), then PCR reaction conditions may need to be modified.

PCR products (usually 5 μL) were mixed with a small volume (1 μL) of load buffer (Appendix A.2), and loaded into sample wells of 2% agarose gels (Appendix A.2). They were then electrophoresed at 10 V cm^{-1} for 50 minutes, conditions which were sufficient to enable estimation of the size of PCR products. Bands of DNA were visualised on a UV transilluminator (Watson Victor UVP Gel Documentation System).

This whole blood PCR protocol did not generate any observable product, except for a few faint bands when all 20 μL of PCR reaction was loaded onto agarose gel (results not shown). The method was continually modified (results not shown) until reproducible results were achieved (see following section).

2.1.4 Revised whole blood PCR

Whole blood PCR was revised so that of 2 μL of whole blood (human subject 1) were mixed with 4 μL water and then microwaved. Mixing with water lysed blood cells, released genomic DNA into the extracellular milieu. PCR reagents were then added as follows:

10 x PCR reaction buffer	2.0 μL
Water	10.8 μL
HucathU (50 mM)	0.4 μL
Bocthex8 (50 mM)	0.4 μL
dNTPs (0.25 mM)	0.2 μL
TAQ polymerase (50 U/ μL)	0.2 μL
Total volume (including 6 μL blood/water)	20.0 μL

This method generated two amplimers differing in size by about 20 bp (380 and 360 bp). Varying the time period of microwave irradiation identified 2.5 - 3 minutes as the ideal length of time for the diluted blood sample to be irradiated (Figure 2.2).

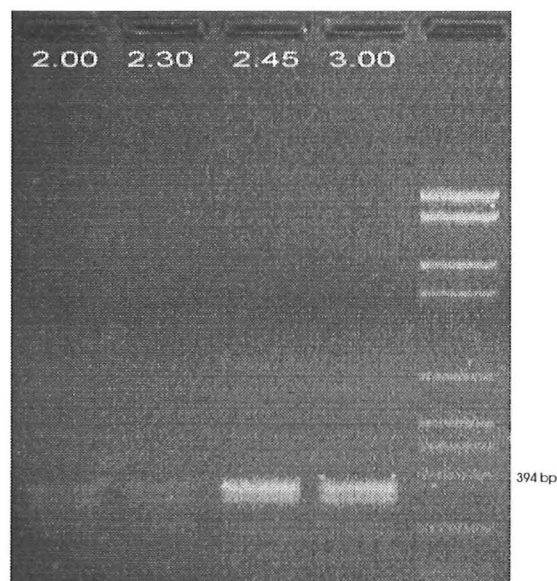


Figure 2.2 Agarose Gel of Whole Blood PCR. Each lane was loaded with 20 μL of PCR product. Contents of each lane (left to right) were PCR reactions containing blood/water mixtures irradiated on high for; 2 minutes, 2 minutes 30 seconds, 2 minutes 45 seconds, and 3 minutes. Lanes 2 and 4 contain two distinct bands of DNA for subject 1. 5 μL of molecular weight marker VI (Appendix A.2) were loaded in lane 5. The 394 bp fragment in the molecular weight marker is labelled.

2.1.5 Discussion

This method of whole blood PCR differed from that of Ohharra *et al.*, (1994), as it did not incorporate the use of Tris-HCl, added KCl, or Triton[®] X-100. Instead, water was used to lyse blood cells by osmotic pressure before microwave irradiation. Two bands of PCR product were produced during PCR amplification.

DNA from subject 4 was available (by consent) through previous work undertaken by Craigen (1995). This DNA was amplified concomitantly with whole blood PCR for comparison, and as a control. Reaction mixtures were prepared as in section 2.1.2 except that 2 μ L of whole blood were substituted with 0.5 μ L of DNA (199 ng/ μ L), and 16.3 μ L water was added to each tube. DNA from subject 4 generated 2 bands of PCR product similar to whole blood PCR results for subject 1, and again was thought to result from non-specific primer binding (results not shown). Steps were then taken to optimise the PCR reaction so that only one amplimer was produced. The whole blood PCR method was abandoned in favour of using preparations of genomic DNA. Further investigation using primers for different genes and blood from different animals, may yet identify this novel method which incorporates water to lyse blood cells, as a quick and easy method for the generation of PCR products.

2.2. Refinement of PCR methodology

2.2.1 Introduction

Non specific primer binding was suspected, because genomic DNA from subjects 1 and 4 generated two amplimers. Attempts to increase binding specificity by raising the annealing temperature and performing a magnesium ion titration with DNA from subject 4 were undertaken.

The magnesium ion titration was performed simultaneously with an increase in the annealing and extension temperature from 60°C to 64°C. Raising the temperature to 64°C meant that primer binding would be more specific, and performing a magnesium ion titration would identify the best ion concentration for optimal Taq polymerase activity. Magnesium ion concentration can also influence the specificity of primer binding.

2 μ L of magnesium chloride solution (0.0, 0.5, 1.0, 1.5, 2.0, 2.5, or 3.0 mM) were added to tubes, and the following reagents added:

DNA (199ng/ μ L)	0.5 μ L
10 x Mg ⁺⁺ free PCR buffer (Boehringer Mannheim)	2.0 μ L
Water	16.3 μ L
HucathU (50 mM)	0.4 μ L
Bocthex8 (50 mM)	0.4 μ L
dNTPs (0.25 mM)	0.2 μ L
TAQ polymerase (50 U/ μ L)	<u>0.2 μL</u>
Total volume (including magnesium chloride)	20.0 μ L

2.2.2 Results

Figure 2.3 illustrates that raising the annealing and extension temperature by 4^oC did not alter primer specificity in any way, because two amplimers were still produced. No amplification was observed at 0.0, and 0.5 mM ion concentrations, but the two amplimers persisted over the range of ion concentrations from 1.0 - 3.0 mM.

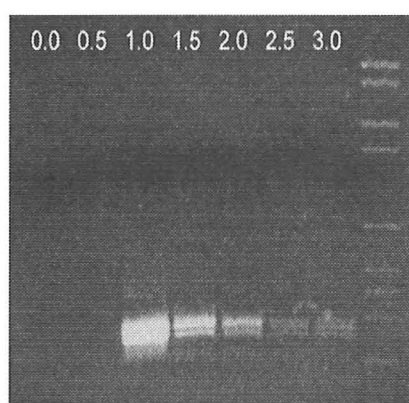


Figure 2.3 PCR banding pattern for subject 4 over a range of magnesium ion concentrations. The ion concentrations (mM) for each reaction tube are listed along the top of the gel in white. 10 μ L of each reaction tube were loaded, and 5 μ L of molecular weight marker VI were loaded in lane 8. (Numbered left to right).

The annealing and extension temperature was decreased to 60^oC, and 2 μ L of 10 x PCR reaction buffer added to all subsequent PCR reactions so that magnesium ion concentrations in each tube were equal to 1.5mM.

Genomic DNA extracted from human colo-rectal tumour tissue was provided by Dr. S. Damak for comparison. This DNA (100ng/ μ L) was amplified along with that of subject 4,

and the water content in each PCR reaction was adjusted to accommodate the differences in DNA concentrations. Agarose gel electrophoresis of PCR products revealed that tumour tissue generated only one amplicon in comparison to the 2 bands for subject 4 (Figure 2.4).

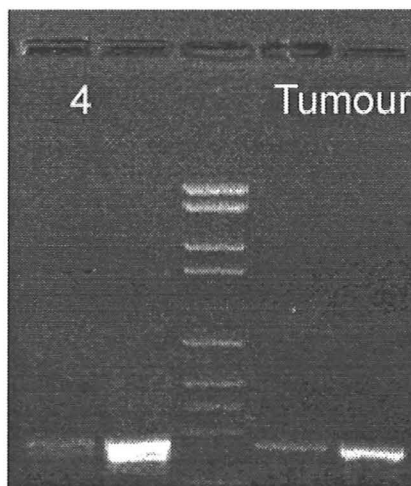


Figure 2.4 PCR banding patterns generated from DNA of human subject 4 and tumour tissue. Lanes 1 and 2 (numbering is from left to right) contained 5 and 15 μ L respectively of PCR product for subject 4. Lanes 4 and 5 contained 5 and 15 μ L of PCR product for human tumour tissue. 5 μ L of molecular weight marker VI were loaded in lane 3.

2.2.3 Discussion

DNA from subjects 1 and 4 generated two bands of PCR product, which were approximately 360 bp and 380 bp. Tumour tissue DNA generated only one PCR product (380 bp). The persistence of two amplicons at the increased annealing temperature over a range of magnesium ion concentrations and the presence of only one amplicon from tumour tissue DNA suggested that more than one allele existed for human cathepsin B. The primers (HucathU and Bocthex8) were perhaps binding both alleles in the heterozygotes (subjects 1 and 4), and the individual from whom the tumour tissue DNA was sourced was possibly homozygous, generating the 380 bp amplicon only.

Different alleles of human cathepsin B gene have not been identified to date, despite the generation of multiple mRNAs from a variety of both normal and transformed tissues (Gong *et al.*, 1993). Results from these experiments were not conclusive for the existence of two alleles, because one DNA sample had been sourced from (abnormal) tumour tissue. To address this issue there was a requirement for genomic DNA to be both sourced and extracted from the same tissue, using a uniform method.

CHAPTER 3

Preparation of 10 human genomic DNA samples for gene analysis

3.1 Introduction

DNA for gene analysis was extracted from human blood. The white blood cell fraction is a primary source of genomic DNA, and although blood-taking is an invasive technique, it is preferred over tissue removal. Approval for human blood to be collected for use in gene analyses was received from the Lincoln University Human Subjects Committee on the condition that volunteers completed the informed consent form in Appendix B.1. The 10 volunteers (subjects 1-10), were current blood donors and assumed to be free of known blood-borne diseases. This distinction was made so that the handling of any blood products before, during, and after DNA extraction was relatively risk free, although all safety precautions pertaining to the use of human biological samples were observed. A follow up of the ten volunteers did not indicate any trauma had arisen as a result of sampling. Human subjects 1 and 4 had participated in previous experiments for this study.

3.2 Hi-Salt DNA extraction from whole blood

The Hi-Salt DNA extraction protocol was obtained from Environmental Sciences and Research (ESR, Canterbury, New Zealand), and is specifically designed for the recovery of DNA from human blood.

10 mL of blood was collected in vacutainer™ blood collection tubes containing EDTA. 200 µL of each blood sample was used for DNA extraction, and the remainder was stored at -18°C. 800 µL of red and white blood cell lysing solution (Appendix B.2) was added to 200 µL of blood in a 1.5 mL eppendorf tube, and each tube gently inverted every five minutes at room temperature for 30 minutes. Samples were then centrifuged at 13000 rpm for 2 minutes (Biofuge 13, Heraeus Sepetech GmbH). The supernatant was discarded and the pellet washed with 100 µL aliquots of washing buffer (pH 8, Appendix B.2), with subsequent centrifugation (13000 rpm for 2 minutes), and repeated disposal of the supernatant. This washing step was repeated either 8 times or until the pellet was white. 400 µL of solubilisation solution (Appendix B.2) was added to the washed pellet and heated for 15 minutes at 100°C. The mixture was then neutralised with 100 µL of pH-8 neutralising buffer (Appendix B.2). Contents of the eppendorf tube were vortexed, centrifuged at 13000 rpm for 2 minutes, and the supernatant containing genomic DNA retained. The

concentration of DNA in each raw sample was then estimated and diluted with water if necessary (section 3.3).

3.3 Estimation of raw genomic DNA concentrations

A blank containing 4 μL of solubilisation buffer, 1 μL of neutralising buffer (pH 8), and 495 μL of water, was placed in the reference beam, and 5 μL of genomic DNA suspension diluted to 500 μL with water (100 x dilution) was placed in the sample path of a double beamed scanning spectrophotometer (Perkin-Elmer). The absorbancies at wavelengths 200-300 nm were recorded. Equation 3.1 was used to calculate the concentration of DNA in the original sample:

Equation 3.1. DNA concentration in $\mu\text{g/mL} = A_{260} \times 50 \times \text{dilution}$

Scanning between these wavelengths also gave an indication of the purity of each sample. A strong peak at 260nm was indicative of a high DNA concentration, and a low peak shoulder at 280nm (wavelength at which the peptide bond absorbs) indicated a sample low in protein. Figure 3.1 represents a scan of a pure DNA sample.

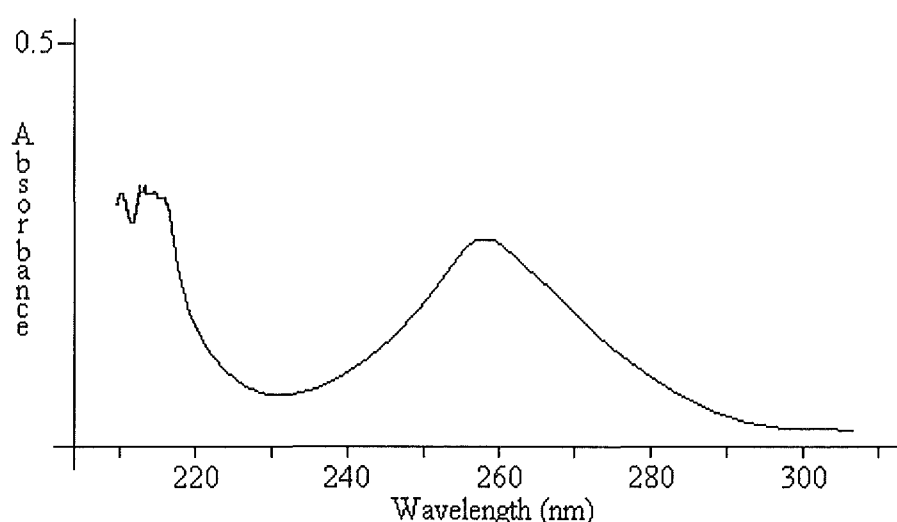


Figure 3.1 Scan of a pure DNA sample (From Couldrey, 1994).

For pure DNA the A_{260}/A_{280} ratio is 1.8, with values less than this indicating the presence of protein and values greater than 1.8 indicating contamination with RNA. Following estimation of concentration appropriate aliquots of each DNA sample were diluted with water to give a concentration of 100 ng/ μL , then stored at 4°C.

3.4 Results and discussion

Table 3.1 lists the DNA concentrations for subjects 1-10 ($\mu\text{g}/\text{mL}$), the total amount of DNA extracted (μg), and the purity of the samples (A_{260}/A_{280}). The A_{260}/A_{280} ratios ranged from 1.01 to 1.15, and indicate that the Hi-Salt method for DNA extraction may generate impure genomic DNA preparations.

Table 3.1 Raw genomic DNA concentrations and purity

Subject	A260	A280	A260 A280	Volume uL	Conc. ug/mL	ug DNA
1	0.043	0.04	1.13	500	215.00	107.50
2	0.173	0.17	1.01	500	865.00	432.50
3	0.110	0.10	1.06	500	550.00	275.00
4	0.051	0.05	1.11	500	255.00	127.50
5	0.050	0.05	1.11	500	250.00	125.00
6	0.046	0.04	1.15	500	230.00	115.00
7	0.071	0.07	1.09	500	355.00	177.50
8	0.045	0.04	1.13	500	225.00	112.50
9	0.049	0.04	1.11	500	245.00	122.50
10	0.048	0.04	1.14	500	240.00	120.00

Unlike many other protocols for DNA extraction from mammalian blood samples the Hi-Salt DNA extraction method does not employ the use of a non-specific protease such as proteinase K, which may explain the high protein contamination. However the protocol is sufficient preparation of genomic DNA "clean" enough for PCR amplification, and it is not especially time-consuming.

CHAPTER 4

PCR amplification of genomic DNA from 10 individuals

4.1 Introduction

PCR was used to produce amplimers from genomic the DNA of 10 human subjects. The banding patterns for subjects 1 and 4 seen in earlier experiments were reproduced, and banding patterns similar to that for DNA extracted from tumour tissue were also observed.

4.2 PCR reagents and reaction protocol

Preparation of 10 human DNA samples of the same concentration (100 ng/ μ L) allowed for PCR reactions to be made up to the following uniform method, which was used in all subsequent PCR reactions:

DNA (100 ng/ μ L)	1.0 μ L
10 x PCR reaction buffer	2.0 μ L
Water	15.8 μ L
HucathU (50 mM)	0.4 μ L
Bocthex8 (50 mM)	0.4 μ L
dNTPs (0.25 mM)	0.2 μ L
TAQ polymerase (50 U/ μ L)	0.2 μ L
Total volume	20.0 μ L

4.3 Results and discussion

Amplification of genomic DNA for subjects 1-10 revealed banding patterns strongly suggestive of two different alleles for human cathepsin B (Figure 4.2).

To confirm that the different banding patterns were a result of different alleles, it was decided that the 360 and 380 bp amplimers should be sequenced. Sequencing the amplimers would confirm that they were produced by amplification of the human cathepsin B gene, and would also characterise the alleles. Subjects 8 and 10 were putative homozygous carriers of each allele, and their corresponding amplimers were chosen for sequencing.

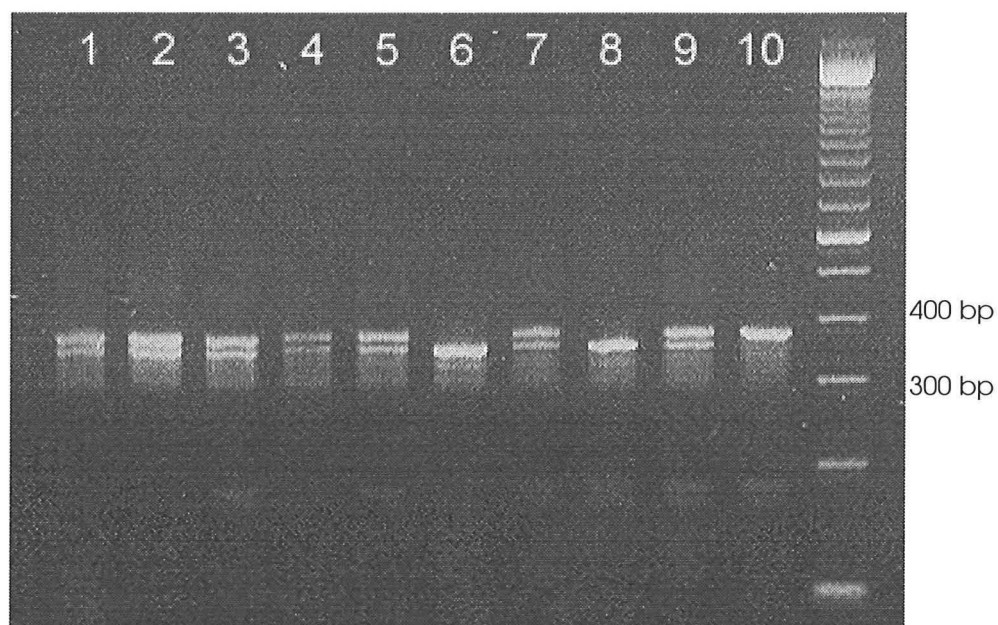


Figure 4.2 Banding patterns suggestive of two different alleles amongst ten human subjects. 10 μ L of each PCR product were loaded, and 5 μ L of 100 base pair ladder (Appendix A.2) were loaded in lane 11. The 300 and 400 bp fragments are indicated (numbering is from left to right).

CHAPTER 5

Preparation of purified amplimers for sequencing

5.1 Introduction

The isolation of PCR products free of any contamination such as protein, genomic DNA, salts, etc. is important, as these impurities can affect DNA sequencing. DNA sequencing required appreciable amounts of template DNA. Promega's Wizard™ PCR Preps DNA Purification System and Boehringer Mannheim's Agarose Gel DNA Extraction Kit were used to purify PCR products for sequencing.

5.1.1 Wizard™ PCR preps DNA purification system (Promega).

70 tubes of PCR product amplified from subject 8's genomic DNA produced 1.4 mL of pooled PCR product. 2% TAE agarose gels (2) were used to separate PCR product from any unused dNTPs etc. After electrophoresis the bands of DNA were cut out from each gel using a sterile scalpel blade, over a UV light source. The 2 excised bands of agarose were then solubilised in 3 mL of purification resin* in a 15 mL Falcon™ tube at 65° C for 6 minutes. 2.0 mL aliquots of the solubilised resin/DNA suspension were poured onto 2 minicolumns using a 3 mL plastic syringe, and each column was then washed with 2 mL of 80% isopropanol. The 2 minicolumns were then placed in a 1.5 mL microcentrifuge tube and centrifuged at 12000 rpm for 20 seconds to dry the resin. They were then transferred to a clean 1.5 mL microcentrifuge tube, and 50 µL of water was pipetted onto the top of each column. After one minute each minicolumn was centrifuged at 12000 rpm for 20 seconds to elute the purified DNA into the 1.5 mL microcentrifuge tube.

PCR product amplified from DNA from subject 10 was purified in the same way.

5.1.2 Agarose gel DNA extraction kit (Boehringer Mannheim).

70 tubes of PCR product amplified from subject 8's genomic DNA produced 1.4 mL of pooled PCR product. 2% TAE agarose gels (2) were used to separate PCR product from any unused

*All reagents used in both methods (except water) were provided with the respective purification kits.

dNTPs etc. After electrophoresis the bands of DNA were cut out from each gel using a sterile scalpel blade, over a UV light source. The excised bands were combined with 12 mL of agarose solubilisation buffer in 15 mL Falcon tube. 10 μ L of silica suspension were then added and the mixture incubated at 56°C for 10 minutes to melt the agarose. During incubation the tubes were vortexed once every 2-3 minutes. The mixture was then centrifuged in an ultracentrifuge (Beckman J2-M1, USA) for 30 seconds at 13000 rpm and the supernatant discarded. The pellet (silica matrix and bound DNA) was resuspended in 500 μ L of nucleic acid binding buffer, and transferred to a 1.5 mL centrifuge tube. This was centrifuged for 30 seconds at 13000 rpm (Biofuge 13, Heraeus Sepetech GmbH), and the supernatant discarded. The pellet was washed twice with 500 μ L of washing buffer with intermittent centrifugation at 13000 rpm for 30 seconds, and the supernatant's discarded. The pellet was then dried over a paper towel at room temperature for 15 minutes, and resuspended in 50 μ L water (pH 8.5), by vortexing every 2-3 minutes over a period of ten minutes. The resulting suspension was then centrifuged for 30 seconds at 13000 rpm, and the supernatant containing purified PCR product kept.

PCR product amplified from DNA from subject 10 was purified in the same way.

5.1.3 Results and discussion

Yields from the Wizard™ PCR prep kit were suspected to be low (results not shown). Recovery for this kit declines with fragment length, with typical yields of 70-90% for 500 bp fragments and 3% for 50 bp fragments. The smallest purified amplicon was ~ 360 bp, and the worst case scenario meant that product recovery was as low as 52%. The Agarose Gel DNA Extraction Kit (Boehringer Mannheim) was trialed as it should give an 80% recovery for a fragment of similar length (370 bp).

Although it worked sufficiently, the Agarose Gel DNA Extraction Kit (Boehringer Mannheim) was not suited to the isolation of large amounts of DNA from agarose. The kit was used only once for the purification of each amplicon from corresponding subjects 8 and 10, before the provided agarose solubilisation buffer became limiting.

Knowing the concentration of template DNA for sequencing was critical, as too much template DNA can inhibit the sequence reaction. The concentrations of each purified amplicon were measured by one of the two following methods: estimation by gel electrophoresis, and measurement of the absorbency at 260 nm.

5.2 Measurement of purified amplimer concentration by gel electrophoresis

Molecular weight marker VI (Boehringer Mannheim) contains 15 fragments: 154, 154, 220, 234, 234, 298, 298, 394, 453, 517, 653, 1033, 1230, 1766, and 2176 bp, and the sum of all the base pairs ($154 + 154 + 220 + \dots + 2176$), is 9814 bp. The amount of DNA for any one fragment can be calculated if the dilution factor for the molecular weight marker is known. For instance, 10 μL of the diluted molecular weight marker VI (Appendix A.2) contains a total 125 ng of DNA, and equation 5.1 gives the amount of DNA for the 394 bp fragment.

Equation 5.1
$$394/9814 \times 125 \text{ ng} = 5.0 \text{ ng of DNA}$$

The 394 bp fragment is similar in size to the purified amplimers, and when purified amplimers were electrophoresed alongside 10 μL of molecular weight marker VI, their concentrations could be estimated according to band intensities. Choosing a fragment of similar size was important because DNA binds ethidium bromide semi-quantitatively, and a different fragment would have given misleading results.

1 μL of load buffer was mixed with 1 μL of purified amplimer and loaded onto a 2% agarose gel alongside a 10 μL of molecular weight marker VI. After gel electrophoresis the intensity of the band for each purified amplimer was compared to the intensity of the 394 bp fragment, and a simple back-calculation (equation 5.2) gave the DNA concentration for each purified amplimer.

Equation 5.2
$$\text{Intensity factor} \times 5 = [\text{DNA}] \text{ of purified amplimer (ng}/\mu\text{L})$$

5.2.1 Results and discussion

Figure 5.1 illustrates a gel loaded with purified amplimer for subjects 8 and 10. Band intensities were best estimated when the gel was still on the UV transilluminator, rather than from a photograph.

The Wizard™ PCR preps kit produced 100 μL of purified amplimer from 70 PCR tubes. 1 μL of each purified amplimer produced bands 6 times as intense as the 394 bp fragment of the molecular weight marker. Therefore the estimated DNA concentration for each purified was 30 ng/ μL . 80 μL of each purified amplimer were then concentrated to 12 μL in a Speed Vac concentrator (Savant) to achieve a concentration of 200 ng/ μL .

Tubes containing 12 μL of each purified amplimer (200 ng/ μL), were then wrapped in Parafilm™ and posted along with 10 μL of each primer (diluted to 0.8 pmol/ μL) to the University of Otago for sequencing.

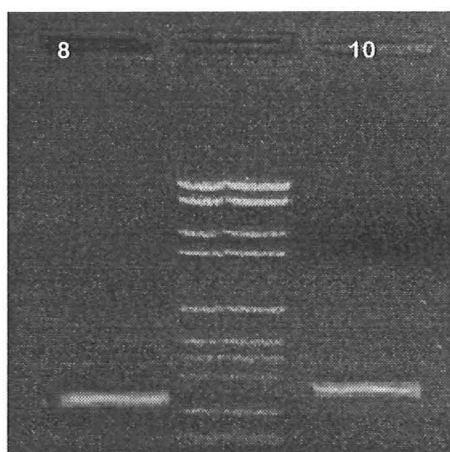


Figure 5.1 Purified amplicons on agarose gel for concentration estimation by band intensity. 1 μ L of amplicon (8 on left, 10 on right) were loaded alongside 10 μ L of molecular weight marker VI.

5.3 Estimation of DNA concentrations using the scanning spectrophotometer

5 μ L of purified amplicon isolated using the Agarose Gel DNA Extraction Kit (Boehringer Mannheim), were diluted to 500 μ L with water, and the absorbency at 260 nm recorded using a double beamed scanning spectrophotometer (Perkin-Elmer). A blank containing 500 μ L of water was placed in the reference beam. DNA concentrations were calculated using equation 5.3.

Equation 5.3 $\text{DNA concentration in } \mu\text{g/mL} = A_{260} \times 50 \times \text{dilution}$

5.3.1 Results and discussion

70 tubes of PCR product produced 50 μ L of purified amplicon for subjects 8 and 10 using the Agarose Gel DNA Extraction Kit (Boehringer Mannheim). The absorbency at 260 nm for amplicon 10 was 0.05450, and 0.07515 for amplicon 8. Therefore the concentration of amplicon 10 was ~ 270ng/ μ L, and ~370 ng/ μ L for amplicon 8. An aliquot of each DNA suspension was then diluted 100 ng/ μ L.

Tubes containing 20 μ L of each purified amplicon (100 ng/ μ L), were then wrapped in Parafilm™ and posted along with 10 μ L of each primer at 3.2 pmol to the University of Auckland for sequencing.

CHAPTER 6

Sequencing results.

6.1 Introduction

Sequencing of amplimers was performed by Tracee Masson-Lawrence of the Centre For Gene Research, University of Otago, New Zealand, and by Debbie Reed of the School of Biological Sciences, University of Auckland, New Zealand. Sequencing of amplimers 8 and 10 at the University of Otago did not produce any meaningful results.

Attempts to sequence amplimers 10 and 8 at the University of Otago failed. Sequencing of the amplimers at the University of Auckland were successful, and revealed each amplimer to be a unique allele for human cathepsin B. The alleles differed due to an extra 19 nucleotides in amplimer 10.

6.2 Sequence failure due to errors in estimation of DNA concentration

Two attempts to sequence amplimers 8 and 10 at the University of Otago failed. Of the sequences received back from this facility, none shared homology with the primers Bocthex8 and HucathU (results not shown), no homology with published human cathepsin B gene sequence could be found (results not shown), and random 20 bp oligomers selected from within sequence for each amplimer did not share homology with any published gene sequences in the NCBI blast server (results not shown).

The Wizard™ PCR Preps purification system (Promega) produced 100 µL of each amplimer at a concentration of 30 ng/µL. Therefore a total of 3000ng of amplimer were purified from the contents of 70 PCR reaction tubes (section 5.2.1). For amplimer 10, the Agarose Gel DNA Extraction Kit (Boehringer Mannheim) produced 50 µL of amplimer at a concentration of 270ng/mL, and a total of 13.5 mg of amplimer were purified from the contents of 70 PCR reaction tubes (section 5.3.1). The differences in recovery rates of these two kits are not great enough to explain the differences between the total amounts of DNA isolated from 70 tubes. The estimation of DNA concentration by comparison of band intensities in agarose gels was concluded to be grossly underestimated. This meant that template DNA sent to the University of Otago was well in excess of 200 ng/µL, and inhibited the sequencing reaction so that no meaningful results were produced.

6.3 Sequence results from the University of Auckland.

Template DNA sent to the University of Auckland produced meaningful sequence results, and identified the measurement of the absorbency at 260 nm as the most reliable method for measuring template DNA concentrations intended for sequencing.

Amplimers 8 and 10 were sequenced using automated dye terminator reactions. The sequence results were good enough to build a consensus sequence for each amplimer (Figure 6.1), using the aligned upstream and downstream primer sequences for each amplimer (Appendix C.1).

Amplimer 10

```
TGTAATGGTG GCTATCCTGC TGAAGCTTGG AACTTCTGGA CAAGAAAAGG 50
CCTGGTTTCT GGTGGCCTCT ATGAATCCCA TGTAGgtaag tgtgtccct 100
tggccacttt ctggccagat ggattgtttg agcaattaac catcatggct 150
ttatttgcct ttataaactg ggggttgaga cagaggggct gctgagaggt 200
gctagccagg tgcacaggcc tctggcagga cctgcctggc gtccatgctg 250
caggcacgag gcttgccttg ccccaggctt gcctcgcca gctggggctt 300
ctccgtgggg gctgtaggtt gactccgctt tctcccgggt cccatcagGG 350
TGCAGACCGT ACTCCATCCC TCCC 374
```

Amplimer 8

```
TGTAATGGTG GCTATCCTGC TGAAGCTTGG AACTTCTGGA CAAGAAAAGG 50
CCTGGTTTCT GGTGGCCTCT ATGAATCCCA TGTAGgtaag tgtgtccct 100
tggccacttt ctggccagat ggattgtttg agcaattaac catcatggct 150
ttatttgcct ttataaactg ggggttgaga cagaggggct gctgagaggt 200
gctagccagg tgcacaggcc tctggcagga cctgcctggc gtccatgctg 250
caggcacgag gcttgccttg ccccaggctt *****t 281
ctccgtgggg gctgtaggtt gactccgctt tctcccgcgt cccatcagGG 331
TGCAGACCGT ACTCCATCCC TCCC 355
```

Figure 6.1 Consensus Sequences for amplimers 10 and 8. The missing 19 nucleotides in amplimer 8 are denoted by "*". g and c denote a possible point mutation. Exon sequence is in upper case, and intron sequence is in lower case.

The first 85 nucleotides in each amplimer share 100% homology with exon 7 of the published human cathepsin B gene sequence (Chan *et al.*, 1986). Amplimer 10 contains an extra 19 nucleotides in intron 7, which explains its reduced electrophoretic mobility observed in agarose gels (eg. Figure 4.2). Only one cathepsin B gene is thought to exist (Gong *et al.*, 1993), therefore these sequencing results identify two unique alleles of this gene. Amplimers 10 and 8 are now referred as alleles A and B respectively.

CHAPTER 7

Discussion of sequence results

7.1 Introduction

Alleles A and B of the human cathepsin B gene differ by 19 nucleotides in intron 7. It is not known if the 19 nucleotides have been inserted into the A allele, or if the 19 nucleotides have been deleted from the B allele. The exact position at which the 19 nucleotides are missing from the B allele cannot be elucidated. This is because the missing 19 nucleotides contain a CT dinucleotide, and the site adjacent to where they are missing from in the B allele contains a CT dinucleotide. Figure 7.1 illustrates this.

```

B      caggcacgag gcttgccttg ccccaggtct ***** *****t 281
B'     caggcacgag gcttgccttg ccccaggt** ***** *****ctt 281
A      caggcacgag gcttgccttg ccccaggtct gcctcgccca gctggggcctt 300
                                   #                               @

```

Figure 7.1 The A allele is aligned with the corresponding region in the B allele missing the 19 nucleotides. B' identifies the alternative region missing the 19 nucleotides in the B allele. CT nucleotides highlighted with # and @ are discussed in text.

Therefore in the B allele the 19 nucleotides are either missing to the left of CT-dinucleotide marked with @ in figure 7.1, or from the right of the CT-dinucleotide marked with a #. The following sections discuss the how the 19 nucleotides may alter cathepsin B gene expression, and the implications of being homozygous for either allele.

7.2 Intron consensus sequences within alleles A and B

Intron 7 of alleles A and B obeyed the GT-AT rule (Lewin, 1994), as well as the general consensus for introns (Figure 7.2), although the 3' 12Py region was not strictly adhered to. Allele A contained three G's at positions 337, 338, and 339, and an A at position 344. Allele B contained two G's at positions 318 and 320, and an A at position 325. The difference in the 12Py regions for the two alleles is discussed in section 7.3.

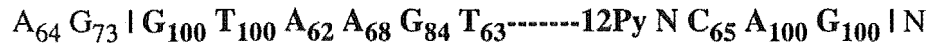


Figure 7.2 General consensus sequence for introns. The **bolded** intron is separated by "|" from the two flanking exons. Subscript numbers indicate the frequencies at which those nucleotides occur (Lewin, 1994).

Gene expression generates primary messenger RNA transcripts containing intron sequences which cannot be translated. Processing of the primary transcript removes the introns and produces secondary mRNA ready for translation. During processing the spliceosome cuts the 5' end of the intron to separate the left exon and the right intron-exon molecule. The 5' end of the intron-exon molecule then forms a lariat through the generation of a 5'-2' bond with the A located in the branch site. Following lariat formation, cutting at the 3' splicing site releases the free intron in lariat form, while the right exon is spliced to the left exon (Lewin, 1994). Failure to process transcripts properly can result in the skipping of exons. The probable branch sites for alleles A and B were located, and aligned with the branch site consensus sequence in Table 7.1.

Table 7.1 Probable branch sites for the A and B alleles. The G at positions 317 and 298 for each allele does not strictly conform to the consensus (Lewin, 1995).

Consensus Sequence	Py80	N	Py80	Py87	Pu75	A	Py95
A allele	G 317	G	T	T	G	A	C
B allele	G 298	G	T	T	G	A	C

7.3 Implications of the extra 19 base pairs in allele A

The OligoTM program was used to identify a stable loop structure (Figure 7.3) which incorporates the extra 19 nucleotides in the A allele. Free energy change for the loop structure was -10.1 kcal mol⁻¹. There was also a stable loop structure formed in the similar region of allele B (Figure 7.4). The free energy change for this loop is -6.8 kcal mol⁻¹.

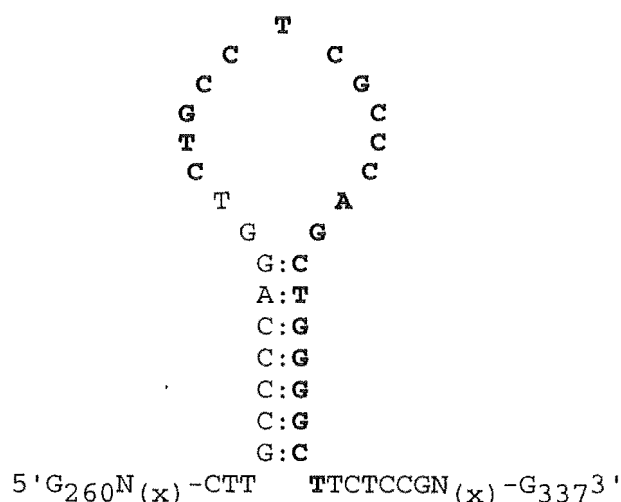


Figure 7.3 The stable loop structure afforded by 19 extra nucleotides (**bolded**) in allele A. The flanking regions started with G at position 260 and ended with the G at position 337.

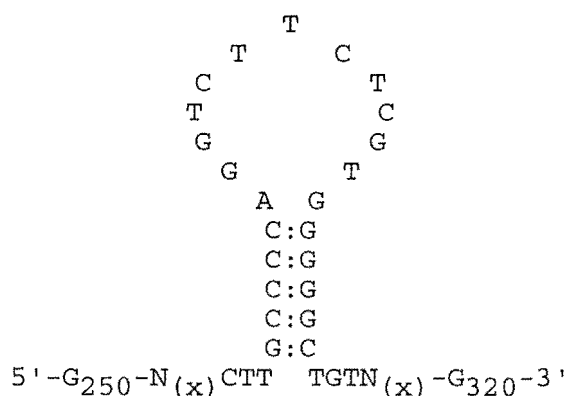


Figure 7.4 The stable loop structure formed in absence of the 19 extra nucleotides in allele B. The flanking regions started with G at position 250 and ended with the G at position 320. The 19 extra nucleotides lied within this region of the A allele.

In the absence of the extra 19 nucleotides, a loop with decreased free energy change in nucleotides 250-320 may form in the primary transcript for allele B. This may confer less stability upon the primary transcript, making it more susceptible to digestion by nucleases. If this occurred *in vivo* then it is possible that the A allele produces primary transcripts with a greater half life than transcripts for the B allele. Therefore individuals homozygous for the A allele (A/A) may display altered cathepsin B gene expression through increased levels of mRNA and protein. Conversely an individual homozygous for the B allele (B/B) may display decreased levels for cathepsin B mRNA and protein. Levels of mRNA and protein for cathepsin B in heterozygotes (A/B) may be intermediate. While one allele in an A/B individual may produce a more stable transcript, this may be offset by concomitant destruction of the less stable transcript.

Alternative branch sites within introns can be used in lariat formation, and in some cases generate altered mRNAs deficient in one or more exons. The extra 19 nucleotides in allele A contained a region which could potentially be used as an alternative branch site (Table 7.2), although it did not strictly conform to the branch site consensus sequence.

Table 7.2 The possible alternative branch site in allele A. The C and G nucleotides flanking the A do not conform to the consensus for the branch site.

Consensus Sequence	Py80	N	Py80	Py87	Pu75	A	Py95
A allele	C 285	G	C	C	C	A	G

Inspection of the downstream sequence indicates that this alternative branch site is not likely to be used. The first alternative down stream splice site (A₃₁₆G) does not conform to the splice site consensus sequence (Figure 7.1), and would destroy the reading frame by adding 31 nucleotides of intron to the 5' end of exon 7.

If the extra 19 nucleotides in allele A have been inserted from another region of the human genome, then their origins are obscure. The 19 nucleotide insert was sent to the NCBI blast server with the CT dinucleotide at either end, but no homology was found with any published gene sequence (search results in appendix C.3).

Current literature detailing mutations in introns, describe those which have detrimental effects on the processing of primary transcripts, to the extent that mutated and prematurely terminated proteins are expressed. Longo *et al.*, (1995) identified one such mutation in a patient suffering from leprechaunism. Using PCR technology they characterised a G \Rightarrow A transition in the first nucleotide in intron 13 of the insulin receptor gene. This point mutation activated a cryptic splice site 27 bp upstream in exon 13, and caused an in-frame deletion of the amino acids 859-867 in the extracellular domain of the β subunit insulin receptor. Sun *et al.*, (1995) describe a similar transition in the 3rd intron for the human low density lipoprotein-receptor gene. This G \Rightarrow A transition results in mRNA lacking exon 3, which codes for the 41 amino acids found in the ligand binding domain of the low density lipoprotein-receptor.

Literature detailing silent mutations within gene intron sequences are scarce, simply because there is more impetus behind identifying and characterising the underlying mutations in genetic disorders that seriously affect health.

7.4 Identification of a possible nucleotide substitution

A nucleotide substitution was identified. Sequencing from the upstream (HucathU) primer revealed a G at position 338 in allele A, and a C at position 319 in allele B. Table 7.3 below illustrates this.

Table 7.3 Possible point mutation between alleles A and B.

Allele						*						
A	T 333	C	C	C	G	G	G	T	C	C	C	A
B	T 314	C	C	C	G	C	G	T	C	C	C	A

Consultation of the signal strengths (Appendix C.2) for these regions supported the existence of the substitution. Figure 7.5 illustrates that in the A allele there were signals for both G and C at position 338, with G being the stronger signal. In allele B there was a lone signal for C at position 319. The absence of a weak underlying G (or any other) signal at position 319 suggests that C is the valid nucleotide at this position in the B allele.

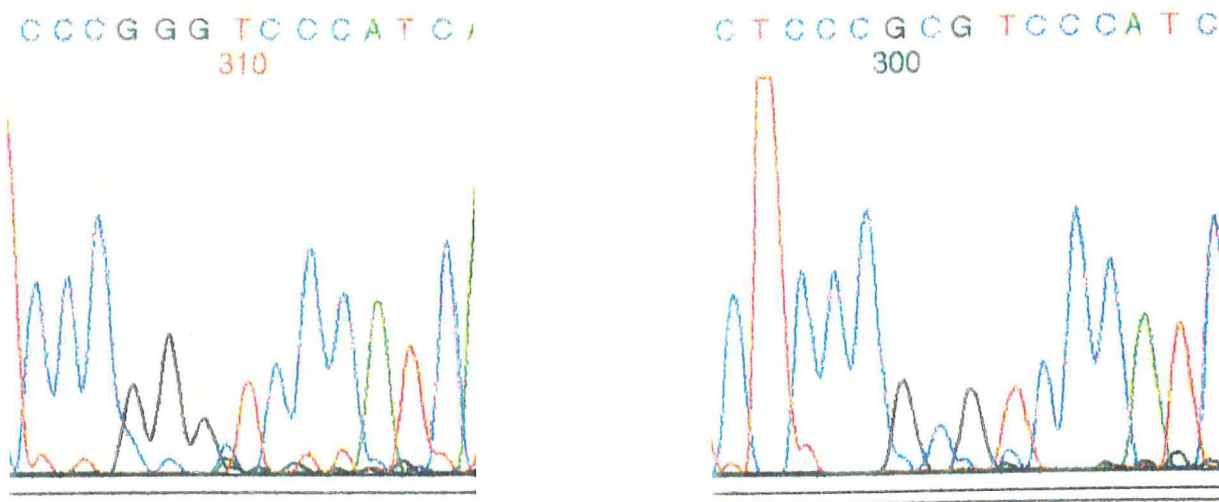


Figure 7.5 Signal strengths for the regions containing possible nucleotide substitution. Allele A is on the left, and allele B is on the right. Position numbers are from signal strengths and should be ignored.

Because these nucleotides reside in the 12 pyrimidine tract of intron 7, the substitution has probably been a C \Rightarrow G substitution at position 338 in allele A. It is less likely to have been a G \Rightarrow C substitution at position 319 in the B allele, as the 12 pyrimidine region is an area generally conserved to pyrimidines only.

It is not known if this substitution is unique to subject 7, or common in all B alleles carried by subjects 1, 2, 3, 4, 5, 6, 7, 8, and 9. Sequencing of the B allele from other homozygous individuals would address this issue. This nucleotide substitution is not be expected to alter processing of the primary mRNA transcript in any way. It serves only to increase the number of purines in the 12 pyrimidine tract in the A allele.

CHAPTER 8

Conclusions

8.1 Conclusions

This study has led to the characterisation of two previously unreported alleles of the human cathepsin B gene. Alleles A and B differ through an extra 19 nucleotides in intron 7 of the A allele. The extra 19 nucleotides in allele A are not thought to affect the processing of primary mRNA transcripts, but can potentially increase the half-life of the transcript. Homozygosity for either allele may lead to altered cathepsin B gene expression, and may result in different cathepsin B to cysteine protease inhibitor ratios in cells. An alteration in this ratio could then enhance tumour development and progression, conditions where cathepsin B gene activity is commonly upregulated.

Alleles A and B also have potential to be used as polymorphic markers in gene linkage analyses of human chromosome 8.

8.2 Future research in this area

The A and B alleles are expected to be lodged in the GENBANK database in late November of 1995. Primers HucathU and Bocthex8 will be used to screen family groups to determine if the alleles are inherited according to Mendelian segregation, and also to observe the frequencies of inheritance for each allele. This work is expected to be undertaken by myself at the John Curtin School of Medical Research in Canberra, Australia, over the summer of 1995-96. Following that, studies on the levels of cathepsin B gene expression in A/A and B/B individuals would be pertinent. It is possible that the presence of primary transcripts for cathepsin B with increased half-lives in transformed cells may promote tumour invasion and progression. Therefore research in this area would be valuable if any associations between tumour development and homozygosity for the A allele are to be found.

References

- Calkins CC, and Sloane BF. 1995 Mammalian cysteine protease inhibitors: biochemical properties and possible roles in tumor progression. *The Journal of Biological Chemistry*. **376**: 71-80.
- Cao L, Taggart RT, Berquin IM, Moin K, Fong D, Sloane B. 1994 Human gastric adenocarcinoma cathepsin B: isolation and sequencing of full length cDNAs and polymorphism's of the gene. *Gene*. **139**: 163-169.
- Chan SJ, San Segundo B, McCormick MB, Steiner DF. 1986 Nucleotide and predicted amino acid sequences of cloned human and mouse preprocathepsin B cDNAs. *Proceedings of the National Academy of Science, U.S.A.* **83(20)**: 7721-7725.
- Chapman HA, Munger JS, Shi G-P. 1994 The role of thiol proteases in tissue injury and remodelling. *American Journal of Respiratory and Critical Care and Medicine*. **150**: S155-S159.
- Craig M-C. 1995 The tissue distribution of ovine glucose transporter RNA : Thesis M Sc, Lincoln University.
- Couldrey C. 1994 Polymorphism in the ovine calpastatin gene. BSc. (Hons) dissertation, Lincoln University, New Zealand.
- Couldrey C. 1995 Personal communication.
- Fong D, Chan MM, Hsieh W-T, Menninger JC, Ward DC. 1992 Confirmation of the human cathepsin B gene (CTSB) assignment to chromosome 8. *Human Genetics*. **89**: 12-12.
- Gong Q, Chan SJ, Bajkowski AS, Steiner DF, Frankfater A. 1993 Characterisation of the Cathepsin B gene and multiple mRNAs in human tissues: Evidence for alternative splicing of Cathepsin B pre-mRNA. *DNA Cell Biology*. **12**: 299-309.
- Lah TT, Kokalj-Kunovar M, Strukelj B, Pungercar J, Barlic-maganja D, Drobic-Korsorok M, Kastelic L, Babnik J, Golouh R, Turk V. 1992 Stefins and lysosomal cathepsins B, L, and D in human breast carcinoma. *International Journal of Cancer*. **50**: 36-44.
- Lewin B. 1994 Stability of mRNA is determined by particular sequences. *Genes V* (Book), Oxford University Press, New York, 1994. pp273-276.
- Longo N, Langley SD, Griffin LD, and Elsas LJ. 1995 Two mutations in the insulin receptor gene of a patient with leprechaunism: application to prenatal diagnosis. *Journal of Clinical Endocrinology and Metabolism*. **81(5)**: 1496-1501.
- Mach L, Mort JS, Glossl J. 1994a Maturation of human procathepsin B. *Journal of Biological Chemistry*. **269(17)**, 13030-13035.

- Mach L, Mort JS, Glossl J. 1994b Noncovalent complexes between the lysosomal proteinase cathepsin B and its propeptide account for stable, extracellular, high molecular mass forms of the enzyme. *Journal of Biological Chemistry*. **269**(17), 13036-13040.
- Makarewicz R, Drewa G, Szymanski W, Skonieczna-Makarewicz I. 1995 Cathepsin B in predicting the extent of the cervix carcinoma. *Neoplasma*. **42**(1): 21-24.
- Markham LJ. 1994 Gene sequence of Ovine cathepsin B: a study of the gene sequence encoding the lysosomal protease cathepsin B. Bsc.(Hons) dissertation, Lincoln University, New Zealand.
- Moin K, Day N, Sameni M, Hasnain S, Hiram T, Sloane B. 1992 Human tumour Cathepsin B. *Journal of Biochemistry*. **285**: 427-434.
- Mordier SB, Bechet DM, Roux M-P, Obled A, Ferrara MJ. 1993 Nucleotide sequence of bovine preprocathepsin B. A study of polymorphism in the protein coding region. *Biochem et Biophysica Acta*. **1174**: 305-311.
- Mordier SB, Bechet DM, Roux M-P, Obled A, Ferrara MJ. 1995 The structure of the bovine cathepsin B gene: genetic variability in the 3' untranslated region. *European Journal of Biochemistry*. **299**: 35-44.
- Morgan JB, Wheeler TL, Koohmaraie M, Crouse M, Savell JD. 1993 Effect of castration on myofibrillar protein turnover, endogenous proteinase activities, and muscle growth in bovine skeletal muscle. *Journal of Animal Science*. **71**(2): 408-414.
- Murnane MJ, Sheahan K, Ozdemirli M, Shuja S. 1991 Stage-specific increases in cathepsin B messenger RNA content in human colorectal carcinoma. *Cancer Research*. **51**: 1137-1142.
- Ohhara M, Kurosu, Esumi M. 1994 Direct PCR of whole blood and hair shafts by microwave treatment. *BioTechniques*. **17**(4): 726-728.
- PCR Protocols. 1990 Edited by MA Innis, DH Gelfand, JJ Sninsky, and TJ White. London: Academic Press.
- Shuja S, Sheahan K, Murnane MJ. 1991 Cysteine endopeptidase activity levels in normal human tissues, colorectal adenomas and carcinomas. *International Journal of Cancer*. **49**: 341-346.
- Sloane BF, Moin K, Sameni M, Tait LR, Rozhin J, Ziegler G. 1994 Membrane association of cathepsin B can be induced by transfection of human breast epithelial cells with c-Ha-ras oncogene. *Journal of Cell Science*. **107**: 373-384.
- Sun X-M, Patel DD, Bhatnagar D, Knight BL, Soutar AK. 1995 Characterisation of a splice-site mutation in the gene for the LDL receptor associated with an unpredictably severe clinical phenotype in English patients with heterozygous FH. *Arteriosclerosis, Thrombosis, and Vascular Biology*. **15**(2): 219-227.
- Tam SW, Cote-Paulino LR, Peak DA, Sheahan K, Murnane MJ. 1994 Human Cathepsin B encoding cDNAs: sequence variations in the 3' untranslated region. *Gene*. **139**: 171-176.

Thein SL, and Wallace RB. 1986 The use of synthetic oligonucleotides as specific hybridization probes in the diagnosis of genetic disorders. *Human Genetics : a practical approach* (ed. KE Davis), p.33-50. IRL Press, Herndon, Virginia.

Turk B, Dolenc I, Zerovnik E, Turk D, Gubensek F, Turk V. 1994 Human cathepsin B is a metastable enzyme stabilised by specific ionic interactions associated with the active site. *Biochemistry*. **33**: 14800-14806.

Zongchao J, Hasnain S, Hiramata T, Lee X, Mort JS, To R, Huber CP. 1995 Crystal structure of recombinant rat cathepsin B and a cathepsin B-inhibitor complex. *The Journal of Biological Chemistry*. **270(10)**: 5527-5533.

Appendix A.1

Primer homology results from the NCBI blast server

Query = **HucathU**
(25 letters)

Database: Non-redundant PDB+GBupdate+GenBank+EMBLupdate+EMBL
575,107 sequences; 395,388,600 total letters.

Sequences producing high-scoring segment pairs:

gb|T72314|T72314 yc68g10.r1 Homo sapiens cDNA clone 858...
gb|R52199|R52199 yg85e07.r1 Homo sapiens cDNA clone 402...
gb|M13230|HUMCTHB Human lysosomal proteinase cathepsin B...
gb|L16510|HUMCATHB Homo sapiens cathepsin B mRNA, complet...
gb|M14221|HUMCTSB Human cathepsin B proteinase mRNA, com...

Homology

>gb|T72314|T72314 yc68g10.r1 Homo sapiens cDNA clone 85890 5' similar to
gb:M14221 CATHEPSIN B PRECURSOR (HUMAN);

Length = 282

Score = 125 (34.5 bits), Expect = 0.0084, P = 0.0083

Identities = 25/25 (100%), Positives = 25/25 (100%), Strand = Plus / Plus

```
Query:      1  TGTAATGGTGGCTATCCTGCTGAAG  25
             |||
Sbjct:    118  TGTAATGGTGGCTATCCTGCTGAAG  142
```

>gb|R52199|R52199 yg85e07.r1 Homo sapiens cDNA clone 40205 5' similar to
gb:M14221 CATHEPSIN B PRECURSOR (HUMAN);

Length = 386

Score = 125 (34.5 bits), Expect = 0.0091, P = 0.0090

Identities = 25/25 (100%), Positives = 25/25 (100%), Strand = Plus / Plus

```
Query:      1  TGTAATGGTGGCTATCCTGCTGAAG  25
             |||
Sbjct:    149  TGTAATGGTGGCTATCCTGCTGAAG  173
```

>gb|M13230|HUMCTHB Human lysosomal proteinase cathepsin B mRNA, 3' end.

Length = 630

Score = 125 (34.5 bits), Expect = 0.0098, P = 0.0098

Identities = 25/25 (100%), Positives = 25/25 (100%), Strand = Plus / Plus

```
Query:      1  TGTAATGGTGGCTATCCTGCTGAAG  25
             |||
Sbjct:     58  TGTAATGGTGGCTATCCTGCTGAAG  82
```

Appendix A.1

Primer homology results from the NCBI blast server

Query = **HucathU**
(25 letters)

Homology (continued)

>gbL16510|HUMCATHB Homo sapiens cathepsin B mRNA, complete cds.

Length = 1996

Score = 125 (34.5 bits), Expect = 0.011, P = 0.011

Identities = 25/25 (100%), Positives = 25/25 (100%), Strand = Plus / Plus

```
Query:      1  TGTAATGGTGGCTATCCTGCTGAAG  25
             |||
Sbjct:    625  TGTAATGGTGGCTATCCTGCTGAAG  649
```

>gbM14221|HUMCTSB Human cathepsin B proteinase mRNA, complete cds.

Length = 2002

Score = 125 (34.5 bits), Expect = 0.011, P = 0.011

Identities = 25/25 (100%), Positives = 25/25 (100%), Strand = Plus / Plus

```
Query:      1  TGTAATGGTGGCTATCCTGCTGAAG  25
             |||
Sbjct:    642  TGTAATGGTGGCTATCCTGCTGAAG  666
```

Appendix A.1

Primer homology results from the NCBI blast server

Query = **BOCTHEX8**

(24 letters)

Database: Non-redundant PDB+GBupdate+GenBank+EMBLupdate+EMBL
575,107 sequences; 395,388,600 total letters.

Sequences producing high-scoring segment pairs:

gb|M13230|HUMCTHB Human lysosomal proteinase **cathepsin B**...
gb|L06075|BOVCATHB Bos taurus **cathepsin B** mRNA, complete ...
gb|L16510|HUMCATHB Homo sapiens **cathepsin B** mRNA, complet...
gb|M14221|HUMCTSB Human **cathepsin B** proteinase mRNA, com...
gb|T72314|T72314 yc68g10.r1 Homo sapiens cDNA clone 858...

Homology

>gb|M13230|HUMCTHB Human lysosomal proteinase cathepsin B mRNA, 3' end.

Length = 630

Score = 120 (33.2 bits), Expect = 0.017, P = 0.017

Identities = 24/24 (100%), Positives = 24/24 (100%), Strand = Minus / Plus

```
Query:      24  TGCAGACCGTACTCCATCCCTCCC  1
              |||
Sbjct:    145  TGCAGACCGTACTCCATCCCTCCC  168
```

>gb|L06075|BOVCATHB Bos taurus cathepsin B mRNA, complete cds.

Length = 1973

Score = 120 (33.2 bits), Expect = 0.019, P = 0.019

Identities = 24/24 (100%), Positives = 24/24 (100%), Strand = Minus / Plus

```
Query:      24  TGCAGACCGTACTCCATCCCTCCC  1
              |||
Sbjct:    535  TGCAGACCGTACTCCATCCCTCCC  558
```

>gb|L16510|HUMCATHB Homo sapiens cathepsin B mRNA, complete cds.

Length = 1996

Score = 120 (33.2 bits), Expect = 0.019, P = 0.019

Identities = 24/24 (100%), Positives = 24/24 (100%), Strand = Minus / Plus

```
Query:      24  TGCAGACCGTACTCCATCCCTCCC  1
              |||
Sbjct:    712  TGCAGACCGTACTCCATCCCTCCC  735
```


Appendix A.2

Visualisation of DNA by gel electrophoresis

Materials:

Water

Water mentioned in text is autoclaved reverse osmosis water.

TAE Stock x 10:

Tris	48.4 g
Glacial acetic acid	11.4 mL
EDTA (0.5M)	<u>20.0 mL</u>
make to 1L, pH = 8	

TBE Stock x 10:

TRIS	108.0 g
Borate	55.0 g
EDTA (Na)	<u>9.3 g</u>
water	make to 1 L

1 x TBE or TAE buffer:

Dilute stock 10 x with ddH₂O.

Add 5µL ethidium bromide (10mg/ml).

Load buffer:

4M Urea	4.8 g	
50% Sucrose	10.0 g	
0.1% Bromophenol Blue	2.0 mL	(0.5% stock)
0.1% Xylenocyanol	2.0 mL	(0.5% stock)
0.25M EDTA	<u>0.6 g</u>	
water	make to 20.0 ml	

Molecular weight markers

100 base pair ladder (GIBCO BRL)

[stock] = 1.0 µg/µL

dilute 10 µL stock and 20 µL load buffer with 170 µL H₂O.

DNA Molecular weight marker VI (Boehringer Mannheim).

[stock] = 0.25µg/µL

dilute 10µL stock and 20µL load buffer with 170µL H₂O.

Appendix A.2

Preparation of a 2% agarose gel.

Method

0.6g Agarose (SeaKem) mixed with 30.0 ml 1 x TBE.

Record weight, heat to dissolve agarose completely (microwave for 2 x 30 seconds on high).

Allow to cool. Adjust weight back to original with H₂O.

Pour into prepared gel mould. Allow to set.

Set up tank with gel and 1 x TBE buffer.

Prepare samples by mixing 2 µL load buffer + 18 µL sample (Or to same proportions).

Load all 20 µL (or 5 µL), and load 5 µL molecular weight marker in an adjacent well.

Electrophorese at 90V for 50 minutes. Visualise bands using UV Transilluminator and photograph results, or continue electrophoreses if needed.

Preparation of a 2% TAE agarose gel.

The same as above, except 1 x TBE is substituted with 1 x TAE buffer.

Appendix B.1

CONSENT TO PARTICIPATE IN A LABORATORY EXPERIMENT.

I, _____ hereby give my consent to participate in a laboratory experiment as part of a study for the subject BICH 460 at Lincoln University.

This experiment will consist of the withdrawal of 10 mL of my blood by venipuncture (a sterile disposable needle inserted into my vein) carried out by a qualified medical technologist, registered nurse or medical practitioner. The blood will be used for DNA extraction, and gene analysis.

The possible risks of this procedure have been fully explained to me. I understand these risks to include the small but finite possibility of subdermal haematoma (bleeding under the skin), infection from bacterial, viral, or other agents, complications of haemophilia (uncontrolled bleeding), and moderate but transient pain, distress, or discomfort.

I understand that the giving of this blood sample will be of no direct benefit to me. I understand that in the event of medical complications arising as a result of this experiment, Lincoln University provides the facilities of the student health service to assist me. I undertake to notify Lincoln University immediately of any complications resulting from my participation.

I further understand that I have the right to refuse to participate in this experiment, and that if I do refuse no adverse reaction or effect on my marks, or my standing as either a student or staff member will result. I acknowledge that I have been granted sufficient time to consult with my family, or legal guardian, medical or otherwise advisors before signing this consent.

Having consideration to all of the foregoing I do hereby freely and voluntarily give my consent.

Signed: _____ Signed: _____

Name: _____ Name: J. R. MacKenzie

Date: _____ Date: / /95

Appendix B.2

Hi-salt DNA extraction reagents.

Red and white blood cell lysing solution.

0.17 M Ammonium chloride (NH_4Cl).

Wash buffer.

10 mM Sodium Chloride (NaCl) 10mM EDTA, pH = 8.

Solubilisation solution.

50 mM Sodium Hydroxide (NaOH).

Neutralisation buffer.

1M Tris-HCL, pH = 8.

Appendix C.1

Homology between A and B alleles

Comparison of 10 DOWNC
with 10 UP
with 8 UP
with 8 DOWNC

```

HucathU      TGTAATGGTGGCTATCCTGCTGAAG
Exon 7       CTGTAATGGTGGCTATCCTGCTGAAGCTTGGAACTTCTGGACAAGAAAAGGCCTGGTTTCTGGTGGCCTC
10 DOWNC    -TGTAATGGTGGCTATCCTGCTGAAGCTTGGAACTTCTGGACAAGAAAAGGCCTGGTTTCTGGTGGCCTC
10 UP       -----GGGAA--ACG-----GGACANGAAAAGGG--TGTTGCTGGTGGCCTC
8 UP        -----TTCGNACACGGGGCCTTCTGGACAAGAAAAGGCCTGGTTTCTGGTGGCCTC
8 DOWNC     -TGTAATGGTGGCTATCCTGCTGAAGCTTGGAACTTCTGGACAAGAAAAGGCCTGGTTTCTGGTGGCCTC
              ^^^          ^^^^          ^          ^^          ^

Exon 7       TATGAATCCCATGTAGGtaag-----Intron7-----
10 DOWNC    TATGAATCCCATGTAGGTAAGTGTGTCCCCTTGGCCACTTTCTGGCCAGATGGATTGTTTGAGCAATTAA
10 UP       TATGAATCCCATGTAGGTAAGTGTGTCCCCTTGGCCACTTTCTGGCCAGATGGATTGTTTGAGCAATTAA
8 UP       TATGAATCCCATGTAGGTAAGTGTGTCCCCTTGGCCACTTTCTGGCCAGATGGATTGTTTGAGCAATTAA
8 DOWNC    TATGAATCCCATGTAGGTAAGTGTGTCCCCTTGGCCACTTTCTGGCCAGATGGATTGTTTGAGCAATTAA

Intron 7     -----
10 DOWNC    CCATCATGGCTTTATTTGCCTTTATAAACTGGGGGTTGAGACAGAGGGGCTGCTGAGAGGTGCTAGCCAG
10 UP       CCATCATGGCTTTATTTGCCTTTATAAACTGGGGGTTGAGACAGAGGGGCTGCTGAGAGGTGCTAGCCAG
8 UP       CCATCATGGCTTTATTTGCCTTTATAAACTGGGGGTTGAGACAGAGGGGCTGCTGAGAGGTGCTAGCCAG
8 DOWNC    CCATCATGGCTTTATTTNCCTTTATAAACTGGGGGTTGAGACAGAGGGGCTGCTGAGAGGTGCTAGCCAG
              ^

Intron 7     -----
10DOWNC    GTGCACAGGCCTCTGGCAGGACCTGCCTGGCGTCCATGCTGCAGGCACGAGGCTTGCCCTTGCCCCAGGTC
10 UP       GTGTACAGGACTCTGGCAGGACCTGTCTGGCGTCCATGCTGCAGGNACGAGGCTTGCCCTNGCCCCAGGTC
8 UP       GTGCACAGGCCTCTGGCAGGACCTGCCTGGCGTCCATGCTGCAGGCACGAGGCTTGCCCTNGCCCCAGGTC
8 DOWNC    GTGCACAGGCCTCTGGCAGGACCTGCCTGGCGTCCATGCTGCAGGCACGAGGCTTGCCCTTGCCCCAGGTC
              ^      ^          ^          ^          ^

Intron 7     -----
10 DOWNC    TCCCTCGCCCAACTGGGGCTTNCCTCCGTGGGGNCTGTANGT-GACCCCGCTNTCTG--CGCCAGTGCTCA
10 UP       TGCCTCGCCCAGNTGGGGCTT-CTCCGTGGGGGCTGTAGGTTGACTCCGCTTTCTC-CCGGGTCCCATCA
8 UP       T-----TCTCCGNGGGGGGTGTAGGTTGACTCCGCTTTCTC-CCGCGTCCCATCA
8 DOWNC    T-----TNCCTCCGTGGGGGCTGTAGGNTGACCCCGCTTTCCTNCCGCANTCGCCCG
              @@@@@@@@@@@@@@@@@@%$      ^      ^      ^      ^      ^      ^      ^      ^
              $      $^

-Bocthex8    TGCAGACCGTACTCCATCCCT-CCC
Exon 8       GGGTGCAGACCGTACTCCATCCCT-CCC
10 DOWNC    A-----
10 UP       GGGTGCAGACCGTACTCCANCCCT-CNC
8 UP       GGGTGCAGACCGTACTCCATCCCTACCC
8 DOWNC    GGNCCCTAA-----
              %

```

Key "^" identifies 1 nucleotide which differs from three others at that position. The corresponding nucleotide in the consensus sequence is based on the agreement between the 3 similar nucleotides. Published nucleotides (Exon 7 and Exon 8) were also used where appropriate.

"@" identifies each of the 19 nucleotide insertion in allele A

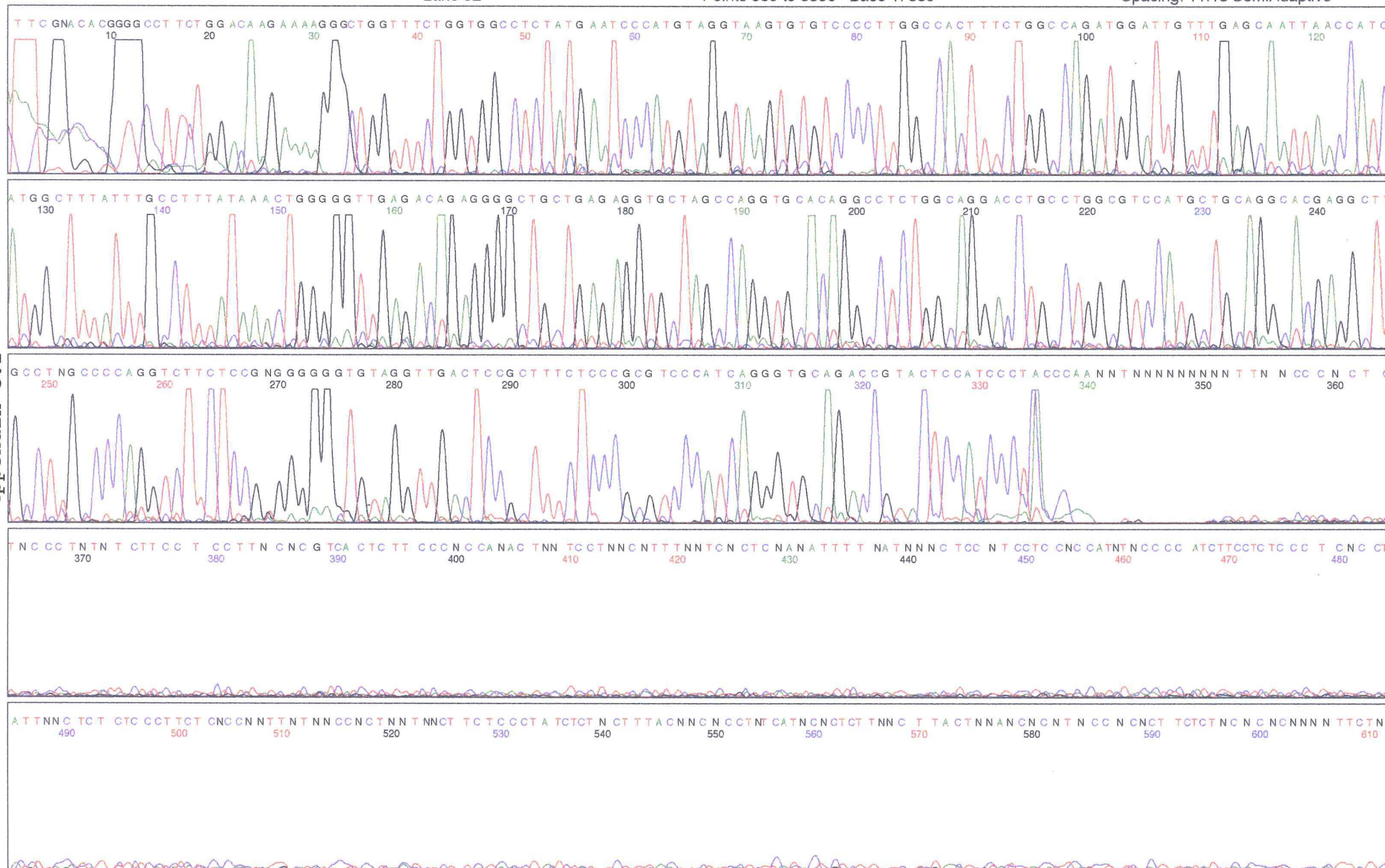
"\$" identifies nucleotides which needed to be resolved from the signal strengths

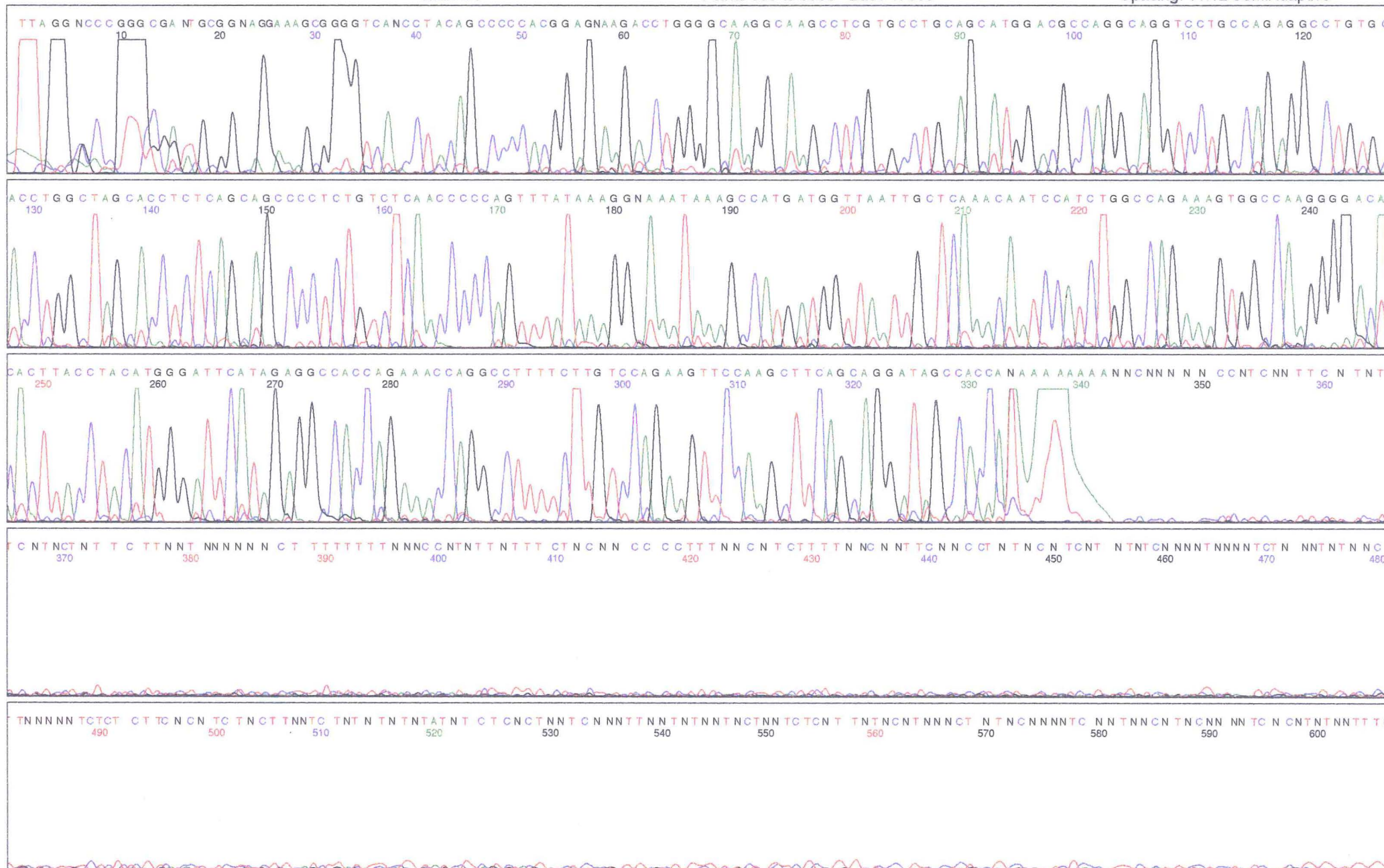
"*" identifies other possible polymorphism between alleles A and B.

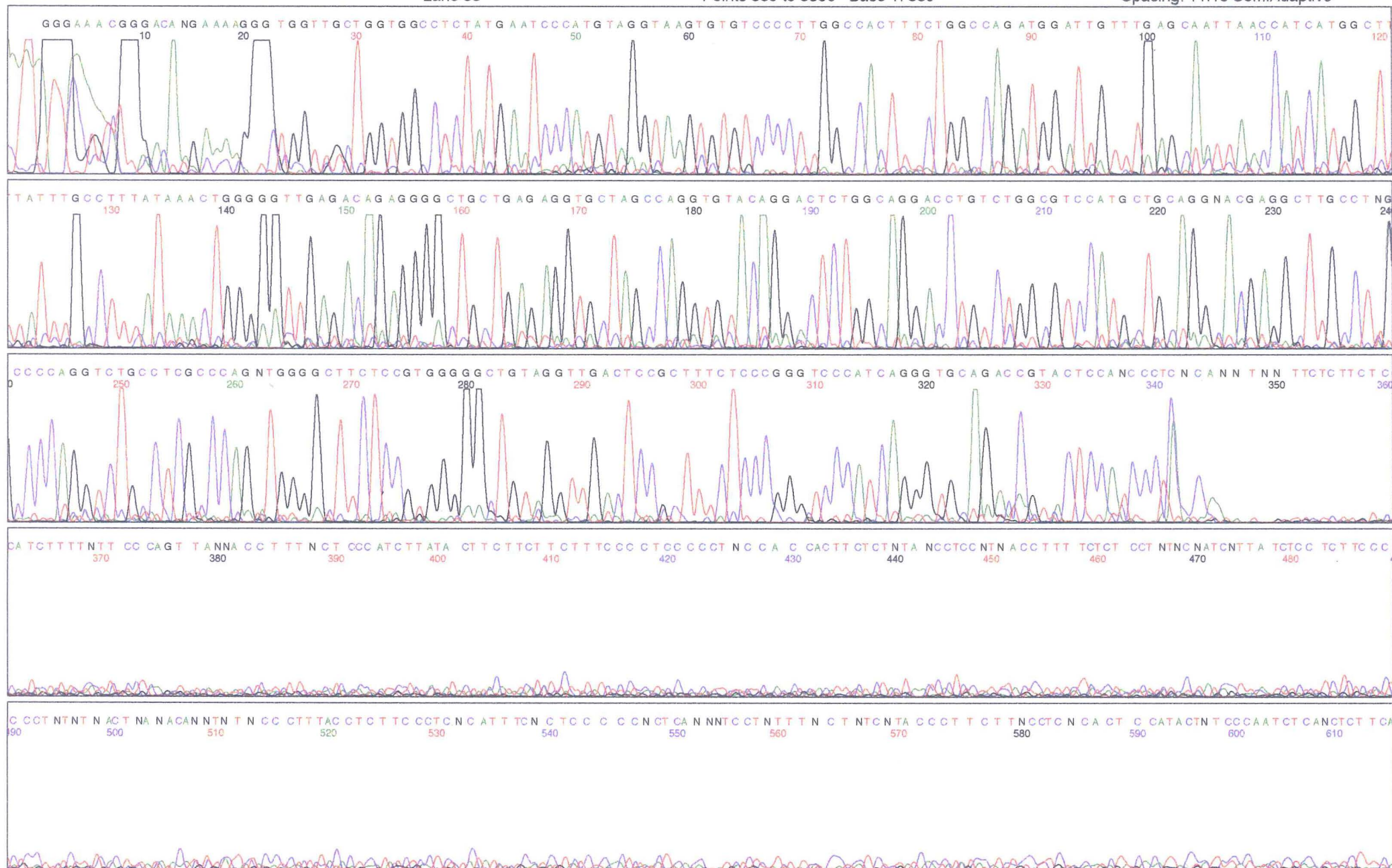
"%" identifies nucleotides thought to be inserted by the gel reader, and which were not included in the consensus sequence.

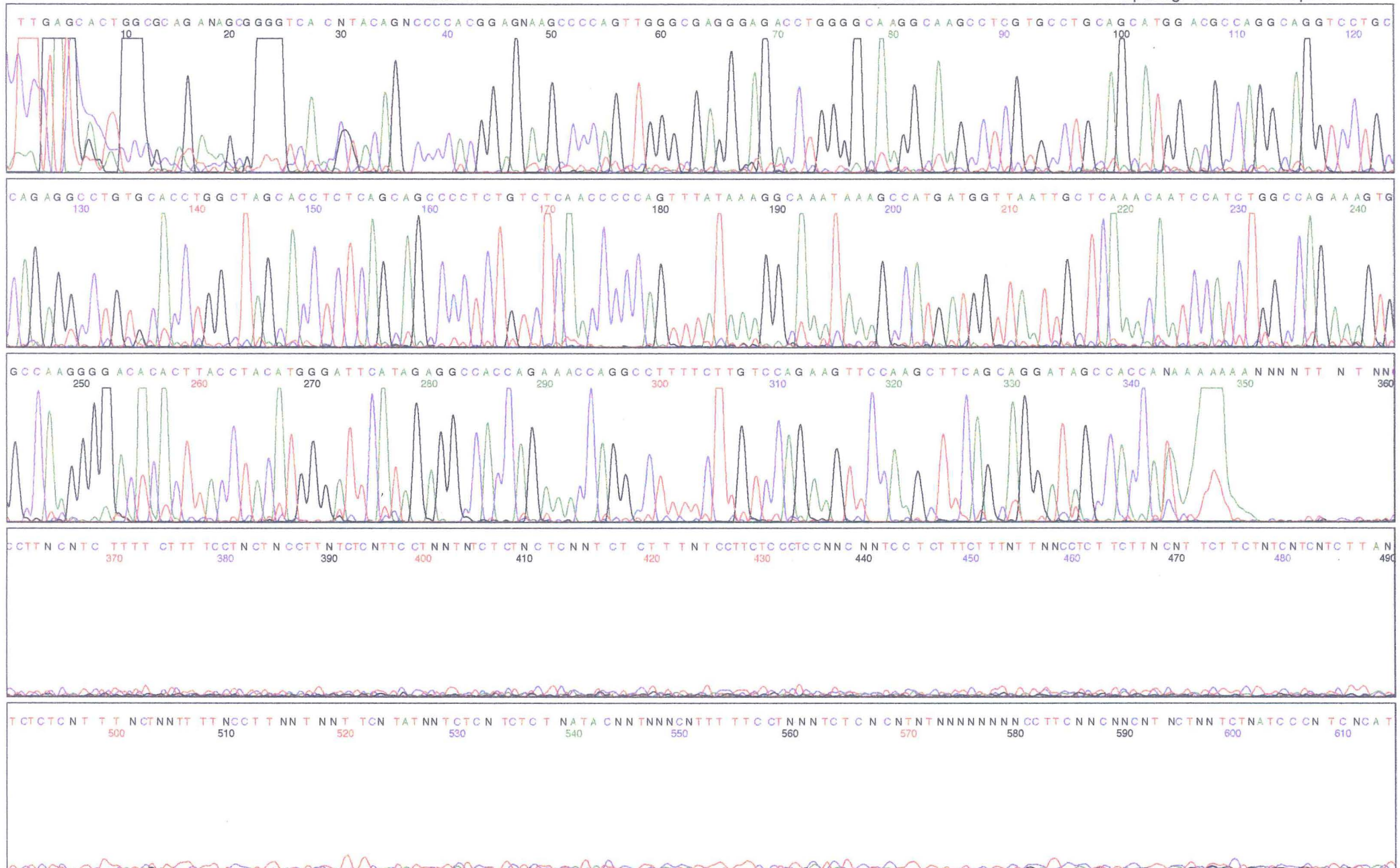
Amplimer from subject 10 represents Allele A

Amplimer from subject 8 represents Allele B









Appendix C.3

19 nucleotide sequence search results

Query = insertct

> gcctcgcccagctggggct

(19 letters)

Database: Non-redundant PDB+GBupdate+GenBank+EMBLupdate+EMBL
575,107 sequences; 395,388,600 total letters.

Sequences producing High-scoring Segment Pairs:	Score	P(N)	N
---	-------	------	---

*** NONE ***

Query = ctinsert

> ctgcctcgcccagctgggg

(19 letters)

Database: Non-redundant PDB+GBupdate+GenBank+EMBLupdate+EMBL
575,107 sequences; 395,388,600 total letters.

Sequences producing High-scoring Segment Pairs:	Score	P(N)	N
---	-------	------	---

*** NONE ***